

# Bayesian modeling of temporal dependence in large sparse contingency tables

TSUYOSHI KUNIHAMA AND DAVID B. DUNSON

*Department of Statistical Science, Duke University, Durham, NC 27708-0251, USA*

tsuyoshi.kunihama@stat.duke.edu

dunson@stat.duke.edu

May, 2012

## Abstract

In many applications, it is of interest to study trends over time in relationships among categorical variables, such as age group, ethnicity, religious affiliation, political party and preference for particular policies. At each time point, a sample of individuals provide responses to a set of questions, with different individuals sampled at each time. In such settings, there tends to be abundant missing data and the variables being measured may change over time. At each time point, one obtains a large sparse contingency table, with the number of cells often much larger than the number of individuals being surveyed. To borrow information across time in modeling large sparse contingency tables, we propose a Bayesian autoregressive tensor factorization approach. The proposed model relies on a probabilistic Parafac factorization of the joint pmf characterizing the categorical data distribution at each time point, with autocorrelation included across times. Efficient computational methods are developed relying on MCMC. The methods are evaluated through simulation examples and applied to social survey data.

*Key words:* Dynamic model; Multivariate categorical data; Nonparametric Bayes; Panel data; Parafac; Probabilistic tensor factorization; Stick-breaking.

## 1 Introduction

Time-indexed multivariate categorical data are collected in many areas, with partially-overlapping categorical variables measured for different subjects at the different time points. As a motivating application, we consider social science surveys that are conducted at regular time intervals, containing many categorical questions such as gender, race, age group, ethnicity, religious affiliation, political party and preference for particular policies. For such surveys and other types of time-indexed multivariate categorical data, it is common for the variables measured (questions asked) to vary somewhat over time while a subset of the variables will be measured at all times. In addition, the number of variables measured can

be moderate to large leading to a contingency table with an *enormous* number of cells, the vast majority of which are empty. Given the fact that social science data often contain complex interactions, it becomes extremely challenging to build realistic and computationally tractable models that allow ultra-sparse data. We define ultra-sparse contingency tables as having exponentially or super-exponentially more cells than the sample size.

Let  $\mathbf{x}_{ti} = (x_{ti1}, \dots, x_{tip})'$  denote the multivariate response for the  $i$ th subject in the survey at time  $t$ , with the  $j$ th categorical question having  $d_j$  elements,  $x_{tij} \in \{1, \dots, d_j\}$ ,  $j = 1, \dots, p$ . We accommodate the case in which the specific variables measured can vary across time by introducing missingness indicators,  $m_{ti} = (m_{ti1}, \dots, m_{tip})'$ , with  $m_{tij} = 1$  if variable  $j$  is missing for subject  $i$  at time  $t$ ; we allow design-based missingness in which certain variables are not measured for any subjects at a particular time and for individual-specific missingness in which certain individuals fail to answer all the questions posed to them. In both cases we assume missing at random.

There is a rich literature on the analysis of contingency tables (Agresti (2002); Fienberg and Rinaldo (2007)). Log linear models are perhaps the most commonly used modeling framework. Routine implementations rely on maximum likelihood estimation, though there is also a rich Bayesian literature. For large, sparse contingency tables, maximum likelihood estimates do not exist in many cases except for overly-simplistic log-linear models and richer classes of models become challenging to implement computationally. There is a rich literature on graphical modeling approaches to estimating conditional independence structures in categorical variables, with Dobra and Lenkoski (2011) proposing a recent Bayesian approach. Although their method is computationally efficient, except for very small tables, the number of possible graphical models is so enormous that it becomes infeasible to visit more than a vanishingly small fraction of the models making accurate model selection or averaging difficult. To facilitate scaling to large tables, Dunson and Xing (2009) and Bhattacharya and Dunson (2011) recently proposed Bayesian probabilistic tensor factorizations. These methods express the probability tensor corresponding to the joint probability mass function of the categorical variables as a convex combination of independent components. Such methods have not yet been developed for time-indexed contingency tables.

There is a rich literature on categorical time series and longitudinal data analysis in

which the same categorical variable is repeatedly measured for each subject over time. For example, Markov models, state space models and random effects models are routinely applied in such settings. However, these models are not relevant to the problem of incorporating dependence over time in modeling of large sparse contingency tables. As different subjects are measured at different times, we are not faced with the problem of incorporating within-subject dependence in repeated observations; instead our goal is to include dependence in the parameters characterizing the time-dependent joint pmfs for the categorical variables. To our knowledge, this problem has not yet been addressed in the literature. Although one can potentially adapt log-linear or graphical models developed for a contingency tables at one time in a somewhat straightforward manner, the hurdles mentioned above for the static case are compounded in the dynamic setting.

To facilitate routine implementations in ultra sparse cases, we propose to adapt the Dunson and Xing (DX) (2009) probabilistic Parafac factorization to the dynamic setting. The DX model induces a tensor factorization through a Dirichlet process (DP) mixture of product multinomial distributions for the categorical observations. There is an increasingly rich literature proposing nonparametric Bayes dynamic models, which allow time-indexed dependent random probability measures. Perhaps the most common approach relies on a dependent DP (MacEachern (1999, 2000)), which incorporates time dependence in the weights and/or atoms in a stick-breaking representation (Griffin and Steel (2006); Rodriguez and Horst (2008); Chung and Dunson (2011)). Most applications of dependent DPs fix the weights and allow the atoms to vary, as varying weights can lead to computational complexities. For dynamic modeling of contingency tables, it is more parsimonious to allow varying weights and varying atoms can lead to a substantial computational burden. An alternative approach, which allows varying weights in a computationally convenient and flexible manner, relies on dynamic mixtures of DPs (Dunson (2006); Ren et al. (2010)). Recently, a class of probit stick-breaking processes was proposed (Chung and Dunson (2009)), which has the appealing feature of allowing one to induce time dependence in random probability measures through Gaussian time series models (Rodriguez and Dunson (2011)).

We propose a new nonparametric state space model for time-indexed ultra sparse contingency tables. Relying on a DX-type probabilistic Parafac factorization, we place a dynamic

model on the weights, which relies on transformed normal random variables in a similar manner to probit stick-breaking. The model is nonparametric in the sense that the induced prior for each time-indexed joint pmf assigns positive probability in arbitrarily small neighborhoods of any “true” data-generating pmf. Hence, our model can allow higher-order interactions and complex dependences, while shrinking towards a low-dimensional structure and borrowing information across time to address the curse of dimensionality. In addition, and crucially for the approach to be useful in the motivating applications, posterior computation can be implemented via a highly efficient Markov chain Monte Carlo (MCMC) algorithm relying on a slice sampler related to Kalli et al. (2011). Finally, the factorization produces a low-dimensional representation of the joint pmf, which is otherwise characterized by a daunting number of parameters in many cases, as the number of cells of the tables can be truly massive.

## 2 Model specification

### 2.1 Modeling of multivariate categorical data

We review the nonparametric Bayes approach of Dunson and Xing (2009) for a static large sparse contingency table. Let  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$  be multivariate categorical data for the  $i$ th subject, with  $x_{ij} \in \{1, \dots, d_j\}$ ,  $j = 1, \dots, p$ . Let

$$\boldsymbol{\pi} = \{\pi_{c_1 \dots c_p}, c_j = 1, \dots, d_j, j = 1, \dots, p\} \in \Pi_{d_1 \dots d_p}$$

be a probability tensor where  $\pi_{c_1 \dots c_p} = P(x_{i1} = c_1, \dots, x_{ip} = c_p)$  is a cell probability and  $\Pi_{d_1 \dots d_p}$  is the set of all probability tensors of size  $d_1 \times \dots \times d_p$ . Dunson and Xing (2009) show that any  $\boldsymbol{\pi} \in \Pi_{d_1 \dots d_p}$  can be decomposed as

$$\boldsymbol{\pi} = \sum_{h=1}^k \nu_h \boldsymbol{\Psi}_h, \quad \boldsymbol{\Psi}_h = \boldsymbol{\psi}_h^{(1)} \otimes \dots \otimes \boldsymbol{\psi}_h^{(p)} \quad (1)$$

where  $\boldsymbol{\nu} = (\nu_1, \dots, \nu_k)'$  is a probability vector,  $\boldsymbol{\Psi}_h \in \Pi_{d_1 \dots d_p}$  and  $\boldsymbol{\psi}_h^{(j)} = (\psi_{h1}^{(j)}, \dots, \psi_{hd_j}^{(j)})'$  is a  $d_j \times 1$  probability vector for  $h = 1, \dots, k$  and  $j = 1, \dots, p$ . This expression relies on a Parafac tensor factorization (Harshman (1970) and Kolda (2001)). It follows that any multivariate

categorical data distribution can be expressed as a mixture of product multinomials,

$$P(x_{i1} = c_1, \dots, x_{ip} = c_p) = \pi_{c_1 \dots c_p} = \sum_{h=1}^k \nu_h \prod_{j=1}^p \psi_{hc_j}^{(j)}.$$

By introducing a latent class index  $s_i \in \{1, \dots, k\}$  for the  $i$ th subject, the multivariate responses  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$  are conditionally independent given  $s_i$ . Instead of conditioning on a fixed  $k$ , Dunson and Xing (2009) developed a nonparametric Bayes approach that lets

$$\begin{aligned} \boldsymbol{\pi} &= \sum_{h=1}^{\infty} \nu_h \boldsymbol{\Psi}_h, \quad \boldsymbol{\Psi}_h = \boldsymbol{\psi}_h^{(1)} \otimes \dots \otimes \boldsymbol{\psi}_h^{(p)}, \\ \boldsymbol{\psi}_h^{(j)} &\sim \text{Dirichlet}(a_{j1}, \dots, a_{jd_j}), \text{ independently for } j = 1, \dots, p, \\ &h = 1, \dots, \infty, \\ \nu_h &= V_h \prod_{l < h} (1 - V_l), \\ V_h &\sim \text{beta}(1, \alpha), \text{ independently for } h = 1, \dots, \infty, \end{aligned} \tag{2}$$

where  $a_{jl} > 0$  for  $l = 1, \dots, d_j$  and  $\alpha > 0$ . Although (2) allows infinitely many components, the number  $k_n$  occupied by the  $n$  subjects in the sample will tend to be  $k_n \ll n$ , so few components will be occupied. The model corresponds to a Dirichlet process mixture of product multinomial distributions relying on a stick-breaking representation (Sethuraman (1994)). A prior is induced on the joint pmf which has large support in the sense of assigning positive probability to  $L_1$  neighborhoods of any true joint pmf.

## 2.2 Modeling of time-indexed multivariate categorical data

Relying on the DX type probabilistic Parafac factorization, we propose a new nonparametric Bayes approach for time-indexed large sparse contingency tables. In a dynamic setting, we obtain the time-indexed multivariate response  $\mathbf{x}_{ti} = (x_{ti1}, \dots, x_{tip})'$ ,  $x_{tij} \in \{1, \dots, d_j\}$ , for the  $i$ th subject at time  $t$  for  $i = 1, \dots, n_t$ ,  $t = 1, \dots, T$  and  $j = 1, \dots, p$ . At time  $t$  we have a probability tensor  $\boldsymbol{\pi}_t$  for the multivariate categorical response given by

$$\boldsymbol{\pi}_t = \{\pi_{tc_1 \dots c_p}, c_j = 1, \dots, d_j, j = 1, \dots, p\} \in \Pi_{d_1 \dots d_p}$$

where  $\pi_{tc_1 \dots c_p} = P(x_{ti1} = c_1, \dots, x_{tip} = c_p)$  is a cell probability at time  $t$ . Relying on the probabilistic Parafac factorization, each probability tensor  $\boldsymbol{\pi}_t$  can be expressed as a mixture

of product multinomials

$$\boldsymbol{\pi}_t = \sum_{h=1}^{k_t} \nu_{th} \boldsymbol{\Psi}_{th}, \quad \boldsymbol{\Psi}_{th} = \boldsymbol{\psi}_{th}^{(1)} \otimes \cdots \otimes \boldsymbol{\psi}_{th}^{(p)} \quad (3)$$

where  $k_t \in \mathbb{N}$ ,  $\boldsymbol{\nu}_t = (\nu_{t1}, \dots, \nu_{tk_t})'$  is a probability vector,  $\boldsymbol{\Psi}_{th} \in \Pi_{d_1 \cdots d_p}$  and  $\boldsymbol{\psi}_{th}^{(j)} = (\psi_{th1}^{(j)}, \dots, \psi_{thd_j}^{(j)})'$  is a  $d_j \times 1$  probability vector for  $h = 1, \dots, k_t$ . Letting  $s_{ti} \in \{1, \dots, k_t\}$  denote a latent class index for the  $i$ th subject at time  $t$ , the observations  $\mathbf{x}_{ti}$  are conditionally independent given  $s_{ti}$ .

To borrow information across time, we place a dynamic structure on the probability tensor  $\boldsymbol{\pi}_t$  in (3) assuming time varying weights  $\nu_{th}$  and static atoms  $\boldsymbol{\psi}_{th}^{(j)} = \boldsymbol{\psi}_h^{(j)}$ . Time dependence is induced in the weights through a state space model, which assumes that stick-breaking increments on  $\nu_{th}$  arise through transforming Gaussian autoregressive processes using a monotone differentiable link function  $g : \mathbb{R} \rightarrow (0, 1)$ . This characterization is motivated by the probit stick-breaking process (Chung and Dunson (2009); Rodriguez and Dunson (2011)), and leads to a parsimonious but flexible characterization of time-dependence in joint pmfs underlying large, sparse contingency tables.

Similarly to expression (2), we develop a nonparametric Bayes approach that sets the number of components to  $k_t = \infty$ , though the number of occupied components will tend to be much less than the sample size and can vary across time. The specific model is

$$\boldsymbol{\pi}_t = \sum_{h=1}^{\infty} \nu_{th} \boldsymbol{\Psi}_h, \quad \boldsymbol{\Psi}_h = \boldsymbol{\psi}_h^{(1)} \otimes \cdots \otimes \boldsymbol{\psi}_h^{(p)}, \quad (4)$$

$$\boldsymbol{\psi}_h^{(j)} \sim \text{Dirichlet}(a_{j1}, \dots, a_{jd_j}), \text{ independently for } j = 1, \dots, p, \quad (5)$$

$$h = 1, \dots, \infty,$$

$$\nu_{th} = g(W_{th}) \prod_{l < h} \{1 - g(W_{tl})\}, \quad (6)$$

$$W_{th} = \alpha_{th} + \varepsilon_{th}, \quad \varepsilon_{th} \sim N(0, \sigma_\varepsilon^2), \quad (7)$$

$$\alpha_{th} = \mu + \phi \alpha_{t-1h} + \eta_{th}, \quad \eta_{th} \sim N(0, \sigma_\eta^2), \quad (8)$$

where  $|\phi| < 1$ ,  $\{\varepsilon_{th}\}$  and  $\{\eta_{th}\}$  are sequences of independently normally distributed random variables with mean 0 and variance  $\sigma_\varepsilon^2$  and  $\sigma_\eta^2$  respectively. The parameter  $\phi$  controls the autocorrelation over time in the weights  $\nu_{th}$  on the different components. For sake of parsimony and simplicity in modeling and computation, we include a single time-stationary correlation parameter  $\phi$  instead of allowing dependence to be time or element specific. In the limiting case in which  $\phi = 0$ , the weights  $\nu_{th}$  will be modeled as independent. This does not mean that independent priors are placed on the unknown joint pmfs at each

time, as the incorporation of common atoms automatically induces some degree of *a priori* dependence. However, in applications one typically expects that the joint pmfs will be quite similar over time, and by using varying weights one does not rule out arbitrarily large changes in the pmfs over time. When  $\phi$  is close to one, there will be very high time dependence in the weights, leading to effective collapsing on a model that assumes a single time stationary joint pmf. For the initial state variables, we assume the stationary distributions,  $\alpha_{1h} \sim N(\mu/(1-\phi), \sigma_\eta^2/(1-\phi^2))$  independently for  $h = 1, \dots, \infty$ . Also, we choose priors  $\mu \sim N(\mu_0, \sigma_0^2)$ ,  $\phi \sim U(-1, 1)$ ,  $\sigma_\varepsilon^2 \sim IG(m_\varepsilon/2, S_\varepsilon/2)$  and  $\sigma_\eta^2 \sim IG(m_\eta/2, S_\eta/2)$  respectively.

Expressions (4)-(8) induce a prior on the time-dependent joint pmfs, but it is not immediately obvious how the chosen hyperpriors in the hierarchical specification impact the properties of the prior for  $\{\pi_t\}$ . In particular, it is important to obtain characterizations of the moments of the induced prior for the cell probabilities, as well as the prior covariance between different elements and across time. Such expressions are provided in Lemma 1, with the proof provided in Appendix A. Lemma 2 shows that the prior is well defined in the sense that  $\sum_{h=1}^{\infty} \nu_{th}$  converges to one almost surely.

Lemma 1. The expectation, variance and covariance of the joint prior on the elements of  $\{\pi_t\}$  induced through (4)-(8) are

$$E\{\pi_{tc_1 \dots c_p}\} = \prod_{j=1}^p \frac{a_{jc_j}}{\hat{a}_j}, \quad V\{\pi_{tc_1 \dots c_p}\} = \left( \prod_{j=1}^p \frac{a_{jc_j}(a_{jc_j} + 1)}{\hat{a}_j(\hat{a}_j + 1)} - \prod_{j=1}^p \frac{a_{jc_j}^2}{\hat{a}_j^2} \right) \left( \frac{\beta_2}{2\beta_1 - \beta_2} \right),$$

$$Cov\{\pi_{tc_1 \dots c_p}, \pi_{t+kc'_1 \dots c'_p}\} = \left( \prod_{j=1}^p \frac{a_{jc_j}\{a_{jc'_j} + 1(c_j = c'_j)\}}{\hat{a}_j(\hat{a}_j + 1)} - \prod_{j=1}^p \frac{a_{jc_j}a_{jc'_j}}{\hat{a}_j^2} \right) \left( \frac{\gamma_k}{2\beta_1 - \gamma_k} \right),$$

where  $\beta_1 = E\{g(W_{th})\}$ ,  $\beta_2 = E\{g^2(W_{th})\}$ ,  $\gamma_k = E\{g(W_{th})g(W_{t+kh})\}$ ,  $\hat{a}_j = \sum_{l=1}^{d_j} a_{jl}$  and  $1(\cdot)$  is an indicator function.

The expectation of cell probabilities can be expressed as the product of expectations of Dirichlet priors for atoms. The variance and covariance are expressed as the product of two terms, the first one is related to atoms and the second one comes from time varying weights. As  $\mu \rightarrow \infty$ , then  $\beta_2/(2\beta_1 - \beta_2) \rightarrow 1$  and  $\gamma_k/(2\beta_1 - \gamma_k) \rightarrow 1$ , and the variance and covariance will be influenced only by atoms. In such a case, the measure corresponding to the stick-breaking process will become a point mass at a random atom almost surely. In addition,  $\beta_1$ ,  $\beta_2$  and  $\gamma_k$  do not depend on time  $t$ , hence all expectation, variance and covariance are independent of  $t$  though the covariance depends on the time difference  $k$ . Also, the covariance between cell probabilities with  $c_j = c'_j$  for all  $j$  is always positive and, on the other hand, those with  $c_j \neq c'_j$  for all  $j$  have negative covariance. In a special case

in which the hyperparameters in the Dirichlet prior are  $a_{j1} = \dots = a_{jd_j} = a$  the variance and covariance is zero in the limit as  $a \rightarrow \infty$ . The proof is in Appendix A.

Lemma 2.  $\sum_{h=1}^{\infty} \nu_{th} = 1$  almost surely.

Lemma 2 is important in showing that the prior is well defined. The proof is in Appendix B.

Our proposed prior setting is parsimonious but highly flexible in the sense that the induced prior assigns positive probability in arbitrarily small neighborhoods of any true data-generating pmf. Let  $\Pi$  denote the space having elements of the form  $\boldsymbol{\pi} = \{\boldsymbol{\pi}_t \in \Pi_{d_1 \dots d_p}, t \in \{1, \dots, T\}\}$ . We show in Theorem 1 that the proposed prior has large support on  $\Pi$ .

*Theorem 1.* Let  $\mathcal{Q}$  denote the prior on  $\Pi$  through the proposed model and  $\mathcal{N}_{\epsilon}(\boldsymbol{\pi}^0)$  denote an  $L_1$  neighborhood around an arbitrary  $\boldsymbol{\pi}^0 \in \Pi$ . Then for any  $\boldsymbol{\pi}^0 \in \Pi$  and  $\epsilon > 0$ , the prior assigns positive probability in the  $\epsilon$ -neighborhood,  $\mathcal{Q}\{\mathcal{N}_{\epsilon}(\boldsymbol{\pi}^0)\} > 0$ .

Since the proposed prior is defined on a space with finitely many components, a straightforward extension of theorem 4.3.1 in Ghosh and Ramamoorthi (2003) ensures that the posterior concentrates in arbitrary small neighborhoods of any true data-generating distribution as the sample size increases.

### 3 MCMC algorithm for posterior computation

For posterior computation in DP mixtures, one common approach is marginalizing out the random probability measure with the Polya urn scheme (Bush and MacEachern (1996)). Avoiding marginalization, Ishwaran and James (2001) developed the blocked Gibbs sampler relying on truncation approximation of the stick-breaking representation. Without truncation, Walker (2007) and Papaspiliopoulos and Roberts (2008) proposed the slice sampler and retrospective MCMC methods respectively. Though the slice sampler is simpler to implement, conditional constraints on sticks can cause slow mixing of the chain. Kalli et al. (2011) proposed a more efficient slice sampler avoiding such a mixing problem.

Relying on a slice sampler related to Kalli et al. (2011), we developed a simple and efficient MCMC algorithm for the proposed model. In the motivating application, we have two types of missing data, design-based missingness and individual-specific missingness. We assume missing at random for both cases and handle the missing data using missingness indicators,  $m_{ti} = (m_{ti1}, \dots, m_{tip})'$ , with  $m_{tij} = 1$  if variable  $j$  is missing for subject  $i$  at time  $t$ . In addition, we introduce latent variables  $u_t = (u_{t1}, \dots, u_{tn_t})'$  for the slice sampler.



The likelihood of  $\{u_t\}$  and  $\{\mathbf{x}_t\}$  given  $\{m_{ti}\}$ ,  $\{\nu_t\}$  and  $\{\psi_h^{(j)}\}$  can be expressed as

$$\prod_{t=1}^T \prod_{i=1}^{n_t} \left\{ \sum_{h=1}^{\infty} 1(u_{ti} < \nu_{th}) \prod_{j: m_{tij}=0} \prod_{l=1}^{d_j} \left( \psi_{hl}^{(j)} \right)^{1(x_{tij}=l)} \right\}.$$

This representation is consistent with the original model setting if latent variables  $\{u_t\}$  are marginalized out. In a special case in which  $g$  is a probit link function, the data augmentation approach in Albert and Chib (2001) can improve efficiency of the posterior sampling by introducing independent normal latent variables  $\{z_{tih}\}$  with mean  $W_{th}$  and variance 1 satisfying

$$P(z_{tih} > 0, z_{til} \leq 0, l < h) = \Phi(W_{th}) \prod_{l < h} \{1 - \Phi(W_{th})\} = \nu_{th} = P(s_{ti} = h).$$

We propose the following MCMC sampling steps:

1. For  $h = 1, \dots, k^*$ , with  $k^* = \max\{s_{ti}\}$ , update  $\psi_h^{(j)}$  from the following Dirichlet full conditional posterior distribution,

$$\text{Dirichlet} \left( a_{j1} + \sum_{(t,i) \in A_{jh}} 1(x_{tij} = 1), \dots, a_{jd_j} + \sum_{(t,i) \in A_{jh}} 1(x_{tij} = d_j) \right).$$

where  $A_{jh} = \{(t, i) : m_{tij} = 0, s_{ti} = h\}$ .

2. Update  $z_{tih}$  from the marginal (w.r.t.  $u_{ti}$ ) conditional posterior distribution,

$$z_{tih} \mid \dots \sim \begin{cases} N_-(W_{th}, 1) & h < s_{ti}, \\ N_+(W_{th}, 1) & h = s_{ti}, \end{cases}$$

where  $N_-(W_{th}, 1)$  and  $N_+(W_{th}, 1)$  denote the normal distributions with mean  $W_{th}$  and variance 1 truncated on  $(-\infty, 0]$  and  $(0, \infty)$  respectively.

3. Update  $W_{th}$  from the normal marginal (w.r.t.  $u_{ti}$ ) conditional posterior distribution,  $N(\hat{W}_{th}, \sigma_{W_{th}}^2)$  where

$$\hat{W}_{th} = \sigma_{W_{th}}^2 \left( \sum_{i: s_{ti} \geq h}^{n_t} z_{tih} + \sigma_{\varepsilon}^{-2} \alpha_{th} \right), \quad \sigma_{W_{th}}^2 = \frac{1}{\sum_{i=1}^{n_t} 1(s_{ti} \geq h) + \sigma_{\varepsilon}^{-2}}.$$

4. Update  $u_{ti}$  from the full conditional distribution,  $\text{Uniform}(0, \nu_{ts_{ti}})$ .

5. Update  $s_{ti}$  from the multinomial full conditional distribution,

$$Pr(s_{ti} = h \mid \dots) = \frac{1(h \in B_{ti}) \prod_{j:m_{tij}=0} \psi_{hx_{tij}}^{(j)}}{\sum_{l \in B_{ti}} \prod_{j:m_{tij}=0} \psi_{lx_{tij}}^{(j)}},$$

where  $B_{ti} = \{h : \nu_{th} > u_{ti}\}$ . To identify the elements in  $\{B_{ti}\}$ , we first update  $\alpha_{th}$  and  $W_{th}$  for  $t = 1, \dots, T$  and  $h = 1, \dots, \tilde{k}$  where  $\tilde{k}$  is the smallest number with  $\sum_{h=1}^{\tilde{k}} \nu_{th} > 1 - \min\{s_{ti}\}$  for all  $t$ .

6. For  $h = 1, \dots, k^*$ , update  $\alpha_{th}$  using the forward filtering backward sampling algorithm by Fr urwirth-Schnatter (1994) and Carter and Kohn (1994), or Kalman filter and the simulation smoother by de Jong and Shephard (1995) and Durbin and Koopman (2002).
7. Update  $\mu$  from the conditional posterior,  $N(\mu_*, \sigma_\mu^2)$  where  $\mu_* = \sigma_\mu^2(\hat{\sigma}^{-2}\hat{\mu} + \sigma_0^{-2}\mu_0)$ ,  $\sigma_\mu^2 = (\hat{\sigma}^{-2} + \sigma_0^{-2})^{-1}$  and

$$\hat{\mu} = \frac{\sum_{h=1}^{k^*} \sum_{t=2}^T (\alpha_{th} - \phi \alpha_{t-1h}) + (1 + \phi) \sum_{h=1}^{k^*} \alpha_{1h}}{k^* \{T - 1 + (1 + \phi)/(1 - \phi)\}}, \quad \hat{\sigma}^2 = \frac{\sigma_\eta^2}{k^* \{T - 1 + (1 + \phi)/(1 - \phi)\}}.$$

8. Update  $\phi$  using the independence MH algorithm in which the proposal distribution is constructed relying on the mode and Hessian of the logarithm of the conditional posterior densities  $\pi(\phi \mid \dots)$ . First, we compute  $\hat{\phi}$  which maximizes (or approximately maximizes) the conditional posterior density. Then, we generated a candidate from a truncated normal distribution  $TN_{(-1,1)}(\phi_*, \sigma_\phi^2)$ , where

$$\phi_* = \hat{\phi} + \sigma_\phi^2 \left. \frac{\partial \log \pi(\phi \mid \dots)}{\partial \phi} \right|_{\phi=\hat{\phi}}, \quad \sigma_\phi^2 = \left\{ - \left. \frac{\partial^2 \log \pi(\phi \mid \dots)}{\partial^2 \phi} \right|_{\phi=\hat{\phi}} \right\}^{-1}.$$

9. Update  $\sigma_\varepsilon^2$  from the conditional distribution,  $IG(\hat{m}_\varepsilon/2, \hat{S}_\varepsilon/2)$  where  $\hat{m}_\varepsilon = Tk^* + m_\varepsilon$  and  $\hat{S}_\varepsilon = \sum_{t=1}^T \sum_{h=1}^{k^*} (W_{th} - \alpha_{th})^2 + S_\varepsilon$ .
10. Update  $\sigma_\eta^2$  from the conditional distribution,  $IG(\hat{m}_\eta/2, \hat{S}_\eta/2)$  where  $\hat{m}_\eta = Tk^* + m_\eta$  and  $\hat{S}_\eta = \sum_{h=1}^{k^*} \sum_{t=2}^T (\alpha_{th} - \mu - \phi \alpha_{t-1h})^2 + (1 - \phi^2) \sum_{h=1}^{k^*} \{\alpha_{1h} - \mu/(1 - \phi)\}^2 + S_\eta$ .

In a case in which  $g$  is another link function, we update  $W_{th}$  using the independent MH algorithm, instead of step 2 and 3 above. We generate a candidate from a normal distribution relying on the mode and Hessian of the logarithm of the conditional posterior densities of  $W_{th}$ .

## 4 Simulation study

In this section, we assess the impact of borrowing of information over time by comparing our proposed method to static approaches, such as Dunson and Xing (DX) (2009), applied separately at each time on simulated data. First, we simulate time-indexed contingency tables from the model shown in expressions (4)-(8) with  $T = 10$ ,  $P = 20$ ,  $d_j = 4$  for all  $j$ ,  $\mu = 0$ ,  $\phi = 0.8$ ,  $\sigma_\varepsilon = 0.1$  and  $\sigma_\eta = 0.8$ . At the respective time points we generated 120, 110, 150, 80, 100, 120, 100, 140, 110 and 150 observations, tiny sample sizes compared with the number of cells. For prior distributions, we assumed  $\psi_h^{(j)} \sim \text{Dirichlet}(1, \dots, 1)$ ,  $\mu \sim N(0, 1)$ ,  $\phi \sim U(-1, 1)$ ,  $\sigma_\varepsilon^2 \sim IG(2.5, 0.025)$ ,  $\sigma_\eta^2 \sim IG(2.5, 0.025)$ . We draw 60,000 MCMC samples after the initial 20,000 samples are discarded as a burn-in period and every fifth sample is saved. We observed that the sample paths were stable and the sample autocorrelations dropped smoothly. Therefore, the chains apparently converged and mixed rapidly.

We first assess performance in estimation of cell probabilities. We picked several cells randomly and report true values, posterior means and 95% credible intervals in Figure 1 (the proposed method) and Figure 2 (DX method). The proposed approach covers all true values in 95% intervals and interval widths are much narrower than for the DX approach consistently across time.

We additionally investigate performance in estimating associations among the categorical variables using the following measure of dependence from Dunson and Xing (2009)

$$\rho_{tjj'}^2 = \frac{1}{\min\{d_j, d_{j'}\} - 1} \sum_{c_j=1}^{d_j} \sum_{c_{j'}=1}^{d_{j'}} \frac{\left(\pi_{tc_j c_{j'}} - \bar{\psi}_{tc_j}^{(j)} \bar{\psi}_{tc_{j'}}^{(j')}\right)^2}{\bar{\psi}_{tc_j}^{(j)} \bar{\psi}_{tc_{j'}}^{(j')}}, \quad (9)$$

where  $\bar{\psi}_{tl}^{(j)} \equiv P(x_{tij} = l) \approx \sum_{h=1}^{k^*} \nu_{th} \psi_{hl}^{(j)}$ . The first row of Figure 3 reports plots of all pairs of true values ( $y$ -axis) and posterior means ( $x$ -axis) of  $\rho_{tjj'}$  at time  $t = 2$  and 7. At each time point, coordinate points by our approach locate closely to the  $y = x$  line, compared to widely scattered points by the DX method. In addition, Table 1 shows correlations between true values and posterior means of  $\rho_{tjj'}$ . Although correlations by the DX method are high, the proposed method consistently produces higher correlations.

	t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10	Total
Proposed	0.948	0.977	0.990	0.977	0.983	0.986	0.985	0.965	0.969	0.968	0.974
DX	0.837	0.794	0.880	0.761	0.766	0.921	0.846	0.817	0.831	0.793	0.841

Table 1: Correlations between true values and posterior means of  $\rho_{tjj'}$  using the first simulation data.

Log linear models provide a standard choice for the analysis of contingency tables.

However, one issue is that flexible log-linear models that accommodate arbitrary interactions among the variables and allow time dependence cannot be applied directly to large, sparse tables. Certainly, maximum likelihood estimates typically do not exist and Bayesian methods that allow an unknown dependence structure do not scale beyond small tables. Dahinden et al. (2010) proposed an approach for high-dimensional log-linear models with interactions, which relies on solving several low-dimensional subproblems that are then combined. An earlier approach by Dahinden et al. (2007) instead relied on L1 penalized log-linear models allowing sparsity of tables. Also, Dahinden et al. (2007) proposed an efficient estimation algorithm for model selection for two level categorical variables.

As a second alternative to our proposed approach, we implemented the method of Dahinden et al. (DH) (2007) in a second simulation example with  $T = 8$ ,  $P = 13$  and  $d_j = 2$  for all  $j$ . Other settings are the same as in the first simulation case. As DH did not consider time-indexed contingency tables, we applied their approach separately at each time point using the logilasso R package, with 5-way cross validation used to choose penalty parameters. The second row of Figure 3 and Table 2 summarize the resulting dependence measures  $\rho_{tjj'}$  at time  $t = 2$  and 7 for each method. For the proposed method, the posterior means are close to true values and correlations between estimates and true values are uniformly high. The DH method has a tendency to underestimate dependence, particularly when true values are low, and has the lowest correlation between the estimates and truth.

	t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	Total
Proposed	0.951	0.978	0.979	0.984	0.986	0.969	0.981	0.944	0.965
DX	0.872	0.803	0.838	0.599	0.807	0.884	0.932	0.827	0.696
DH	0.705	0.557	0.733	0.466	0.725	0.506	0.763	0.487	0.562

Table 2: Correlations between true values and posterior means of  $\rho_{tjj'}$  using the second simulation data.

Finally, to gauge robustness we also simulated data from a time-dependent log-linear model in which all the coefficients of the main effects and interactions between two variables independently follow random walk processes with variance 1 and other higher interactions are zero. The third row of Figure 3 and Table 3 report the estimation results. Although we find less difference among them in this case, the proposed method still shows the best performance.

	t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	Total
Proposed	0.725	0.827	0.768	0.798	0.818	0.916	0.791	0.807	0.817
DX	0.642	0.640	0.726	0.664	0.611	0.864	0.769	0.713	0.724
DH	0.371	0.716	0.821	0.491	0.611	0.877	0.764	0.715	0.624

Table 3: Correlations between true values and posterior means of  $\rho_{tjj'}$  using the third simulation data.

## 5 Analysis of social survey data

In this section, we apply the proposed method to data from the General Social Survey (GSS, <http://www3.norc.ox.ac.uk/GSS+Website>). Our focus is on studying associations among demographic and preference variables over time. We select  $p = 29$  categorical variables from 1994 to 2010, including gender, ethnicity, preference for particular policies and many more listed in the supplemental materials. The GSS was conducted every two years across this time period. The numbers of observations are 2,992 (1994), 2,904 (1996), 2,832 (1998), 2,817 (2000), 2,765 (2002), 2,812 (2004), 4,510 (2006), 2,023 (2008) and 2,044 (2010) respectively. There are abundant missing data in which only a subset of the variables were recorded for an individual, and compared to the number of cells, the sample size is quite small at each time point.

We first compared our proposed approach to log-linear models. Unfortunately, current methodology for fitting log-linear models that allow flexible dependence structures cannot accommodate these data due to the large sparse structure, time variation and abundant missing data. Hence, in order to provide a comparison, we initially focused on a bivariate subset of the data consisting of religious preference ( $i = 1, \dots, 5$ ) and attitude towards abortion ( $j = 1, 2$ ) from 1994 to 2010. We consider the following log-linear Poisson models.

$$\text{Model 1:} \quad N_{tij} \sim \text{Poisson}(N_t \mu_{ij}), \quad \log \mu_{ij} = \lambda + \lambda_i^R + \lambda_j^A + \lambda_{ij}^{RA},$$

where  $N_{tij}$  is count of the cell  $ij$  at time  $t$ ,  $N_t = \sum_i \sum_j N_{tij}$ ,  $\lambda_i^R$  is an effect of the first variable (religious preference),  $\lambda_j^A$  is an effect of the second variable (view of abortion) and  $\lambda_{ij}^{RA}$  is an association term. For identifiability, we assume constraints  $\lambda_5^R = \lambda_2^A = \lambda_{5j}^{RA} = \lambda_{i2}^{RA} = 0$ . Model 1 assumes no time-dependence in cell probabilities  $\mu_{ij} / \sum_{i'} \sum_{j'} \mu_{i'j'}$ .

$$\begin{aligned} \text{Model 2:} \quad N_{tij} &\sim \text{Poisson}(N_t \mu_{tij}), \quad \log \mu_{tij} = \lambda_t + \lambda_{ti}^R + \lambda_{tj}^A + \lambda_{tij}^{RA}, \\ \beta_t &= (\lambda_t, \lambda_{t1}^R, \dots, \lambda_{t4}^R, \lambda_{t1}^A, \lambda_{t11}^{RA}, \dots, \lambda_{t41}^{RA})', \\ \beta_{tl} &= \mu_l + \phi_l \beta_{t-1l} + \varepsilon_{tl}, \quad \varepsilon_{tl} \sim N(0, \sigma_l^2), \quad \text{independently for } l = 1, \dots, 10, \end{aligned}$$

where  $\lambda_{ti}^R$ ,  $\lambda_{tj}^A$  and  $\lambda_{tij}^{RA}$  are effects of the first variable, the second variable and interactions at time  $t$  respectively. We assume  $\lambda_{t5}^R = \lambda_{t2}^A = \lambda_{t5j}^{RA} = \lambda_{ti2}^{RA} = 0$  at each time point and  $\beta_0 = \mathbf{0}$  for the initial values. Model 2 is a time dependent hierarchical model where all parameters in the log-linear model follow AR(1) process independently.

We firstly estimate all models using the data from 1994 to 2008. Then, relying on the estimated parameters, we predict the contingency table in 2010 (Table 4). For the proposed model, we used the same MCMC settings as in the simulation study. For log-linear models,

we estimated parameters using an MCMC algorithm where missing values are imputed from conditional probabilities given observed data at each iteration. For example, we generate the religious preference  $i$  given the view of abortion  $j$  with probability  $\mu_{tij} / \sum_{i'} \mu_{ti'j}$ . For priors, we assumed  $\beta = (\lambda, \lambda_1^R, \dots, \lambda_4^R, \lambda_1^A, \lambda_{11}^{RA}, \dots, \lambda_{41}^{RA})' \sim N(\mathbf{0}, I)$  for Model 1,  $\mu_l \sim N(0, 1)$ ,  $\phi_l \sim U(-1, 1)$  and  $\sigma_l^2 \sim IG(2.5, 0.025)$  for all  $l$  for Model 2. Using Gibbs sampling, we generated posterior samples of  $\mu_l$  and  $\sigma_j^2$  from normal and Inverse-Gamma distributions respectively. For  $\beta$ ,  $\phi_l$ ,  $\beta_t$ , we used a MH algorithm in which candidates were generated from normal distributions relying on the mode and Hessian of the logarithm of the conditional posterior densities. We generated 10,000 MCMC samples after the 1,000 burn-in for Model 1 and 20,000 MCMC samples after the 2,000 burn-in for Model 2 and, for both cases, every fifth sample was saved.

We generated replications at every fifth MCMC iteration and computed average of the following predictive criteria,

$$\begin{aligned} \text{Absolute deviation (AD):} & \quad \sum_{i=1}^5 \sum_{j=1}^2 \left| N_{ij}^{rep} - N_{ij}^{obs} \right|, \\ \text{Mean absolute percentage error (MAPE):} & \quad \frac{1}{10} \sum_{i=1}^5 \sum_{j=1}^2 \left| \frac{N_{ij}^{rep} - N_{ij}^{obs}}{N_{ij}^{obs}} \right|, \end{aligned}$$

where  $N_{ij}^{rep}$  and  $N_{ij}^{obs}$  are the replication and observation of count of the cell  $ij$  respectively. To keep the same total number of replications among all methods, predictions are generated from cell probabilities  $\mu_{ij} / \sum_{i'} \sum_{j'} \mu_{i'j'}$  for Model 1 and  $\mu_{2010ij} / \sum_{i'} \sum_{j'} \mu_{2010i'j'}$  for Model 2. Table 5 reports the prediction results. Although Model 2 produces better performance than Model 1 by incorporating time-dependence, the proposed method clearly outperforms log-linear models in terms of both predictive criteria.

	Protestant	Catholic	Jewish	None	Other	Total
Agree	216	103	21	137	60	537
Disagree	372	182	7	81	47	689
Total	588	285	28	218	107	1226

Table 4: Contingency table of the religious preference and view of abortion in 2010.

	Proposed	Model 1	Model 2
AD	194.4	208.6	204.5
MAPE	0.216	0.232	0.227

Table 5: Prediction results.

Next, we apply the proposed method to all 29 categorical variables. We generated 30,000 MCMC samples after the initial 10,000 samples are discarded as the burn-in and every fifth

sample are saved. We observed the sample paths are stable and the sample autocorrelations are small. Table 6 shows the estimation result of parameters in the time dependent stick-breaking processes. Concerning the measure of time dependence  $\phi$ , the posterior mean is close to 1 and the 95% credible interval locates near 1, which means the weights of the stick-breaking processes have strong time dependence over time.

Parameter	Mean	Stdev.	95% interval
$\mu$	-0.012	0.004	[-0.023, -0.005]
$\phi$	0.988	0.004	[0.978, 0.994]
$\sigma_\varepsilon$	0.062	0.009	[0.046, 0.082]
$\sigma_\eta$	0.126	0.011	[0.104, 0.149]

Table 6: Estimation result of parameters in the proposed stick-breaking process.

Then, we investigate cross interactions among the variables over time. Figure 4 show the posterior means of  $\rho_{tjj'}$  for all pairs in 2002 and 2010. Additional results for other years are included in the supplemental materials. We find the structure of interactions is complex at each time point. Also, though each interaction gradually changes over time, all tables look similar to one another, implying they have close dependence. This is consistent with the result of the strong dependent weights in the stick-breaking processes. Some categorical variables such as Race [ $j = 3$ ], Attitude toward abortion [6], Political party affiliation [9] and Think of self as liberal or conservative [14] intricately correlate with many other variables. On the other hand, zodiac [11] shows little interactions with all other variables. Among all pairs of variables, {Age [1], Marital status [10]}, {Attitude toward abortion [6], Attitude toward homosexual [16]} and {Attitude toward homosexual [16], Attitude toward Marijuana [19]} show strong interactions in the whole period. Also, we observed several pairs of variables showing relatively close interactions over time, such as {Attitude toward abortion [6], Think of self as liberal or conservative [14]}, {Race [3], Political party affiliation [9]} and {Marital status [10], Having gun [17]}. In addition, the views of government expense show moderate interactions, especially to the environment [23], nation's health [24], halting the rising crime [25], dealing with drug addiction [26] and education system [27].

Next, we study trends of dependence between categorical variables. Figure 5 reports the posterior means and 95% credible intervals of  $\rho_{tjj'}$  for pairs with close interactions. We observed various patterns of time paths. For {Age, Marital status}, the interaction increased around 2000 then declined sharply to a lower level. {Race, Political party affiliation} and {Race, Having gun} have peaks in 2006 and the interactions have steeply decreased after that. In addition, we can see similar trends in {Attitude toward abortion, Think of self as lib or con}, {Attitude toward abortion, Attitude toward homosexual}, {Attitude toward homosexual, Attitude toward Marijuana}, {Religion, Attitude toward abortion} and

{Religion, Attitude toward Marijuana}. The interactions have roughly increased over time, especially in the 2000s. On the other hand, the dependence in {Race, Death penalty for murder} decreased at first and kept stable in the middle of the period then declined again. {Having gun, Family income} gradually increased over the period but the difference is small. For {Marital status, Having gun}, the interaction dropped in the middle of the period but recovered recently at the same level as the beginning.

## 6 Discussion

We have demonstrated that the proposed approach is useful in analyzing time-indexed large sparse contingency tables. One interesting extension is to accommodate joint modeling of mixed scale variables consisting of not only categorical data but also continuous and count variables. In such a case, one can potentially model the observed data vector for the  $i$ th subject at time  $t$ ,  $y_{ti} = (y_{ti1}, \dots, y_{tip})'$ , as conditionally independent given latent class variables  $x_{ti} = (x_{ti1}, \dots, x_{tip})'$ , with  $x_{ti}$  modeled exactly as proposed in this article. For example, consider the simple case in which  $p = 2$  with  $y_{ti1} \in \mathbb{R}$  continuous and  $y_{ti2} \in \{1, \dots, d_2\}$  categorical. Then, one can let  $y_{ti1} \sim N(\mu_{x_{ti1}}, \sigma_{x_{ti1}}^2)$  and  $y_{ti2} = x_{ti2}$ , with the proposed probabilistic tensor factorization approach flexibly accommodating dependence in  $y_{ti1}$  and  $y_{ti2}$  through dependence in  $x_{ti1}$  and  $x_{ti2}$ . The induced marginal distribution for the continuous variable  $y_{ti1}$  will be a mixture of normals, with the probability weight on each component potentially varying with the categorical variable  $y_{ti2}$ . This same strategy can be generalized to more complex settings involving many categorical, count, continuous and even functional observations.

Another interesting direction in terms of generalizations is to accommodate dependence in the observations; for example, one may collected multivariate categorical longitudinal data in which the same variables are measured repeatedly on the sample study subjects or the data may have a nested structure. Log linear and logistic regression-type models can be easily generalized to such settings, but clearly encounter computational challenges in large sparse settings. Potentially the simplex factor model of Bhattacharya and Dunson (2012) can be generalized to accommodate such dependence structures through the latent factors, with some challenges arising in terms of developing computationally efficient implementations and models that are both flexible and interpretable.

## Acknowledgement

This work was supported by Nakajima Foundation and grant number R01 ES017240 from the National Institute of Environmental Health Sciences (NIEHS) of the National Institutes



of Heath (NIH). The computational results are mainly generated using Ox (Doornik (2006)).

## A Proof of Lemma 1

The expectation of cell probability is

$$\begin{aligned} E\{\pi_{tc_1 \dots c_p}\} &= E\left\{\sum_{h=1}^{\infty} \nu_{th} \prod_{j=1}^p \psi_{hc_j}^{(j)}\right\} = \sum_{h=1}^{\infty} \left[ E\{\nu_{th}\} \prod_{j=1}^p E\{\psi_{hc_j}^{(j)}\} \right], \\ &= \prod_{j=1}^p E\{\psi_{hc_j}^{(j)}\} \sum_{h=1}^{\infty} E\{\nu_{th}\} = \prod_{j=1}^p E\{\psi_{hc_j}^{(j)}\} = \prod_{j=1}^p \frac{a_{jc_j}}{\hat{a}_j}. \end{aligned}$$

The marginal distribution of  $W_{th}$  can be expressed as  $N(\mu/(1-\phi), \sigma_\eta^2/(1-\phi^2) + \sigma_\varepsilon^2)$ , independent of  $t$  and  $h$ . Hence, we set  $\beta_1 = E\{g(W_{th})\}$  and  $\beta_2 = E\{g^2(W_{th})\}$ . The second moment of cell probability is

$$\begin{aligned} E\{\pi_{tc_1 \dots c_p}^2\} &= E\left[\left\{\sum_{h=1}^{\infty} \nu_{th} \prod_{j=1}^p \psi_{hc_j}^{(j)}\right\} \left\{\sum_{l=1}^{\infty} \nu_{tl} \prod_{j=1}^p \psi_{lc_j}^{(j)}\right\}\right], \\ &= \sum_{h=1}^{\infty} \sum_{l=1}^{\infty} E\{\nu_{th} \nu_{tl}\} E\left\{\prod_{j=1}^p \psi_{hc_j}^{(j)} \psi_{lc_j}^{(j)}\right\}, \\ &= \left[\prod_{j=1}^p E\left\{\left(\psi_{hc_j}^{(j)}\right)^2\right\} - \prod_{j=1}^p E^2\left\{\psi_{hc_j}^{(j)}\right\}\right] \sum_{h=1}^{\infty} E\{\nu_{th}^2\} + \prod_{j=1}^p E^2\left\{\psi_{hc_j}^{(j)}\right\} \sum_{h=1}^{\infty} \sum_{l=1}^{\infty} E\{\nu_{th} \nu_{tl}\}, \\ &= \left(\prod_{j=1}^p \frac{a_{jc_j}(a_{jc_j} + 1)}{\hat{a}_j(\hat{a}_j + 1)} - \prod_{j=1}^p \frac{a_{jc_j}^2}{\hat{a}_j^2}\right) \sum_{h=1}^{\infty} E\{\nu_{th}^2\} + \prod_{j=1}^p \frac{a_{jc_j}^2}{\hat{a}_j^2}, \end{aligned}$$

where

$$\begin{aligned} \sum_{h=1}^{\infty} E\{\nu_{th}^2\} &= \sum_{h=1}^{\infty} E\left[g^2(W_{th}) \prod_{l < h} \{1 - g(W_{tl})\}^2\right], \\ &= \sum_{h=1}^{\infty} \beta_2 \{1 - 2\beta_1 + \beta_2\}^{h-1}, \\ &= \frac{\beta_2}{2\beta_1 - \beta_2}. \end{aligned}$$

Hence,

$$V\{\pi_{tc_1 \dots c_p}\} = \left( \prod_{j=1}^p \frac{a_{jc_j}(a_{jc_j} + 1)}{\hat{a}_j(\hat{a}_j + 1)} - \prod_{j=1}^p \frac{a_{jc_j}^2}{\hat{a}_j^2} \right) \left( \frac{\beta_2}{2\beta_1 - \beta_2} \right). \quad (10)$$

Similarly,

$$\begin{aligned} E\{\pi_{tc_1 \dots c_p} \pi_{t+kc'_1 \dots c'_p}\} &= E \left[ \left\{ \sum_{h=1}^{\infty} \nu_{th} \prod_{j=1}^p \psi_{hc_j}^{(j)} \right\} \left\{ \sum_{l=1}^{\infty} \nu_{t+kl} \prod_{i=1}^p \psi_{lc'_i}^{(i)} \right\} \right], \\ &= \left[ \prod_{j=1}^p E \left\{ \psi_{hc_j}^{(j)} \psi_{hc'_j}^{(j)} \right\} - \prod_{j=1}^p E \left\{ \psi_{hc_j}^{(j)} \right\} E \left\{ \psi_{lc'_j}^{(j)} \right\} \right] \sum_{h=1}^{\infty} E\{\nu_{th} \nu_{t+kh}\} + \prod_{j=1}^p E \left\{ \psi_{hc_j}^{(j)} \right\} E \left\{ \psi_{lc'_j}^{(j)} \right\}, \\ &= \left( \prod_{j=1}^p \frac{a_{jc_j} \{a_{jc'_j} + 1(c_j = c'_j)\}}{\hat{a}_j(\hat{a}_j + 1)} - \prod_{j=1}^p \frac{a_{jc_j} a_{jc'_j}}{\hat{a}_j^2} \right) \sum_{h=1}^{\infty} E\{\nu_{th} \nu_{t+kh}\} + \prod_{j=1}^p \frac{a_{jc_j} a_{jc'_j}}{\hat{a}_j^2}, \end{aligned}$$

where

$$\begin{aligned} E\{\nu_{th} \nu_{t+kh}\} &= E \left\{ \left[ g(W_{th}) \prod_{l < h} \{1 - g(W_{tl})\} \right] \left[ g(W_{t+kh}) \prod_{l < h} \{1 - g(W_{t+kl})\} \right] \right\}, \\ &= E \{g(W_{th})g(W_{t+kh})\} \prod_{l < h} E \{[1 - g(W_{tl})]\{1 - g(W_{t+kl})\}\}, \\ &= E \{g(W_{th})g(W_{t+kh})\} \prod_{l < h} [1 - 2\beta_1 + E\{g(W_{tl})g(W_{t+kl})\}]. \end{aligned}$$

From (7) and (8),  $E \{g(W_{th})g(W_{t+kh})\}$  can be expressed as

$$\begin{aligned} E \{g(W_{th})g(W_{t+kh})\} &= E \{g(\alpha_{th} + \varepsilon_{th})g(\alpha_{t+kh} + \varepsilon_{t+kh})\}, \\ &= E \left\{ g(\alpha_{th} + \varepsilon_{th}) g \left( \frac{1 - \phi^k}{1 - \phi} \mu + \phi^k \alpha_{th} + \sum_{i=0}^{k-1} \phi^i w_{t+k-ih} + \varepsilon_{t+kh} \right) \right\}. \end{aligned}$$

Since  $\alpha_{th}$ ,  $\varepsilon_{th}$ ,  $w_{t+k-ih}$  ( $i = 0, \dots, k-1$ ) and  $\varepsilon_{t+kh}$  are independent of one another and their distributions do not depend on  $t$  or  $h$ , hence  $\gamma_k \equiv E \{g(W_{th})g(W_{t+kh})\}$  is dependent on time difference  $k$  but independent of time  $t$ .

In addition,

$$\begin{aligned} \sum_{h=1}^{\infty} E\{\nu_{th} \nu_{t+kh}\} &= \sum_{h=1}^{\infty} \gamma_k \prod_{l < h} \{1 - 2\beta_1 + \gamma_k\}, \\ &= \frac{\gamma_k}{2\beta_1 - \gamma_k}. \end{aligned}$$

Hence,

$$Cov\{\pi_{tc_1 \dots c_p}, \pi_{t+kc'_1 \dots c'_p}\} = \left( \prod_{j=1}^p \frac{a_{jc_j} \{a_{jc'_j} + 1(c_j = c'_j)\}}{\hat{a}_j(\hat{a}_j + 1)} - \prod_{j=1}^p \frac{a_{jc_j} a_{jc'_j}}{\hat{a}_j^2} \right) \left( \frac{\gamma_k}{2\beta_1 - \gamma_k} \right).$$

Since  $\beta_2/(2\beta_1 - \beta_2) > 0$ ,  $\gamma_k/(2\beta_1 - \gamma_k) > 0$  and (10), cell probabilities with  $c_j = c'_j$  for all  $j$  have positive covariance and, on the other hand, those with  $c_j \neq c'_j$  for all  $j$  have negative covariance.

In a case where  $a_{j1} = \dots = a_{jc_j} = a$ , the variance and covariance are expressed as

$$V\{\pi_{tc_1 \dots c_p}\} = \left( \prod_{j=1}^p \frac{1 + 1/a}{d_j^2 + d_j/a} - \prod_{j=1}^p \frac{1}{d_j^2} \right) \left( \frac{\beta_2}{2\beta_1 - \beta_2} \right),$$

$$Cov\{\pi_{tc_1 \dots c_p}, \pi_{t+kc'_1 \dots c'_p}\} = \left( \prod_{j=1}^p \frac{1 + 1(c_j = c'_j)/a}{d_j^2 + d_j/a} - \prod_{j=1}^p \frac{1}{d_j^2} \right) \left( \frac{\gamma_k}{2\beta_1 - \gamma_k} \right).$$

Hence,  $V\{\pi_{tc_1 \dots c_p}\} \rightarrow 0$  and  $Cov\{\pi_{tc_1 \dots c_p}, \pi_{t+kc'_1 \dots c'_p}\} \rightarrow 0$  as  $a \rightarrow \infty$ .

## B Proof of Lemma 2

To prove  $\sum_{h=1}^{\infty} \nu_{th} = 1$  a.s., it is enough to show  $\sum_{h=1}^{\infty} E\{\log(1-g(W_{th}))\} = -\infty$  (Ishwaran and James (2001)).  $g$  is a non-negative monotone increasing link function:  $\Re \rightarrow (0, 1)$ , therefore  $0 < \beta_1 = E\{g(W_{th})\} < 1$ . Then, using Jensen's inequality,

$$E[\log\{1 - g(W_{th})\}] \leq \log[1 - E\{g(W_{th})\}] = \log(1 - \beta_1) < 0.$$

Therefore,  $\sum_{h=1}^{\infty} E\{\log(1 - g(W_{th}))\} = -\infty$  at each time point.

## C Proof of theorem

The proposed prior probability assigned to  $\mathcal{N}_\epsilon(\boldsymbol{\pi}^0)$  can be expressed as

$$\mathcal{Q}\{\mathcal{N}_\epsilon(\boldsymbol{\pi}^0)\} = \int 1(\|\boldsymbol{\pi} - \boldsymbol{\pi}^0\| < \epsilon) d\mathcal{Q}(\boldsymbol{\nu}_t, \boldsymbol{\psi}_h^{(j)}, t \in \{1, \dots, T\}, h = 1, \dots, \infty, j = 1, \dots, p).$$

where  $\boldsymbol{\nu}_t$  is a probability vector induced by the proposed stick breaking process and we use the  $L_1$  distance

$$\|\boldsymbol{\pi} - \boldsymbol{\pi}^0\| = \sum_{t=1}^T p_t \sum_{c_1=1}^{d_1} \cdots \sum_{c_p=1}^{d_p} |\pi_{tc_1 \cdots c_p} - \pi_{tc_1 \cdots c_p}^0|,$$

where  $p_t$  is a probability mass function for time  $t \in \{1, \dots, T\}$ .

For any  $\boldsymbol{\pi}^0 \in \Pi$ , each component in  $\boldsymbol{\pi}^0$  can be expressed as

$$\boldsymbol{\pi}_t^0 = \sum_{h=1}^{k_t} \nu_{th}^0 \Psi_{th}, \quad \Psi_{th} = \boldsymbol{\psi}_{th}^{(1)} \otimes \cdots \otimes \boldsymbol{\psi}_{th}^{(p)},$$

where  $k_t \in \mathbb{N}$ ,  $\boldsymbol{\nu}_t^0 = (\nu_{t1}^0, \dots, \nu_{tk_t}^0)'$  is a probability vector,  $\Psi_{th} \in \Pi_{d_1 \cdots d_p}$  and  $\boldsymbol{\psi}_{th}^{(j)} = (\psi_{th1}^{(j)}, \dots, \psi_{thd_j}^{(j)})'$  is a  $d_j \times 1$  probability vector. We define  $k_0^+ = 0$  and  $k_t^+ = \sum_{i=1}^t k_i$  for  $t = 1, \dots, T$ . Then, we construct  $\boldsymbol{\pi} = \{\boldsymbol{\pi}_t, t \in \{1, \dots, T\}\} \in \Pi$  induced by the proposed prior such that the component with the index  $h$  in  $\boldsymbol{\pi}_t^0$  is approximated by the component with the index  $k_{t-1}^+ + h$  in  $\boldsymbol{\pi}_t$ . Let  $\tilde{\boldsymbol{\nu}}_t = (\tilde{\nu}_{t1}, \tilde{\nu}_{t2}, \dots)'$  be a probability vector where  $\tilde{\nu}_{tm} = \nu_{tm-k_{t-1}^+}^0$  for  $k_{t-1}^+ < m \leq k_t^+$  and  $\tilde{\nu}_{tm} = 0$  otherwise, i.e.,  $\tilde{\nu}_{tk_{t-1}^+ + h} = \nu_{th}^0$  for  $1 \leq h \leq k_t$ . For any  $\epsilon$ , we define a set  $D(\boldsymbol{\pi}^0, \epsilon) \subset \Pi$  such that for any  $\boldsymbol{\pi} \in D(\boldsymbol{\pi}^0, \epsilon)$ , each  $\boldsymbol{\pi}_t$  can be expressed as (4) satisfying  $\boldsymbol{\nu} \in \mathcal{N}_{\epsilon'}(\tilde{\boldsymbol{\nu}})$ , where  $\boldsymbol{\nu} = \{\boldsymbol{\nu}_t, t \in \{1, \dots, T\}\}$ ,  $\tilde{\boldsymbol{\nu}} = \{\tilde{\boldsymbol{\nu}}_t, t \in \{1, \dots, T\}\}$  and  $\epsilon' = \epsilon/2 \prod_{j=1}^p d_j$ , and  $\boldsymbol{\psi}_{k_{t-1}^+ + h}^{(j)} \in \mathcal{N}_{\epsilon''}(\boldsymbol{\psi}_{th}^{(j)})$  for  $h = 1, \dots, k_t$  and  $t = 1, \dots, T$  where  $\epsilon'' = \epsilon/2 \sum_t p_t k_t \prod_j d_j$ .

We consider the intervals  $(a_{th}, b_{th})$  in the real line for  $W_{th}$  in the proposed prior for  $h = 1, \dots, k_t^+$  and  $t = 1, \dots, T$  where

$$a_{th} = \begin{cases} g^{-1}\{\max(\tilde{\nu}_{th} - \tilde{\epsilon}, 0)\}, & (h = 1), \\ g^{-1}\left\{\frac{\max(\tilde{\nu}_{th} - \tilde{\epsilon}, 0)}{\prod_{l < h}\{1 - g(W_{tl})\}}\right\}, & (h = 2, \dots, k_t^+), \end{cases} \quad b_{th} = \begin{cases} g^{-1}\{\tilde{\nu}_{th} + \tilde{\epsilon}\}, & (h = 1), \\ g^{-1}\left\{\frac{\tilde{\nu}_{th} + \tilde{\epsilon}}{\prod_{l < h}\{1 - g(W_{tl})\}}\right\}, & (h = 2, \dots, k_t^+), \end{cases}$$

where  $\tilde{\epsilon} = \epsilon'/2 \sum_t p_t k_t^+$ . In this case, it is straightforward to check  $|\nu_{th} - \tilde{\nu}_{th}| < \tilde{\epsilon}$  for  $h = 1, \dots, k_t^+$  and the proposed prior assigns positive probability to these intervals. Then,

the distance between  $\boldsymbol{\nu}$  and  $\tilde{\boldsymbol{\nu}}$  is

$$\begin{aligned}
\|\boldsymbol{\nu} - \tilde{\boldsymbol{\nu}}\| &= \sum_{t=1}^T p_t \sum_{h=1}^{\infty} |\nu_{th} - \tilde{\nu}_{th}|, \\
&= \sum_{t=1}^T p_t \sum_{h=1}^{k_t^+} |\nu_{th} - \tilde{\nu}_{th}| + \sum_{t=1}^T p_t \sum_{h>k_t^+} \nu_{th}, \\
&< 2\tilde{\epsilon} \sum_{t=1}^T p_t k_t^+ = \epsilon'.
\end{aligned} \tag{11}$$

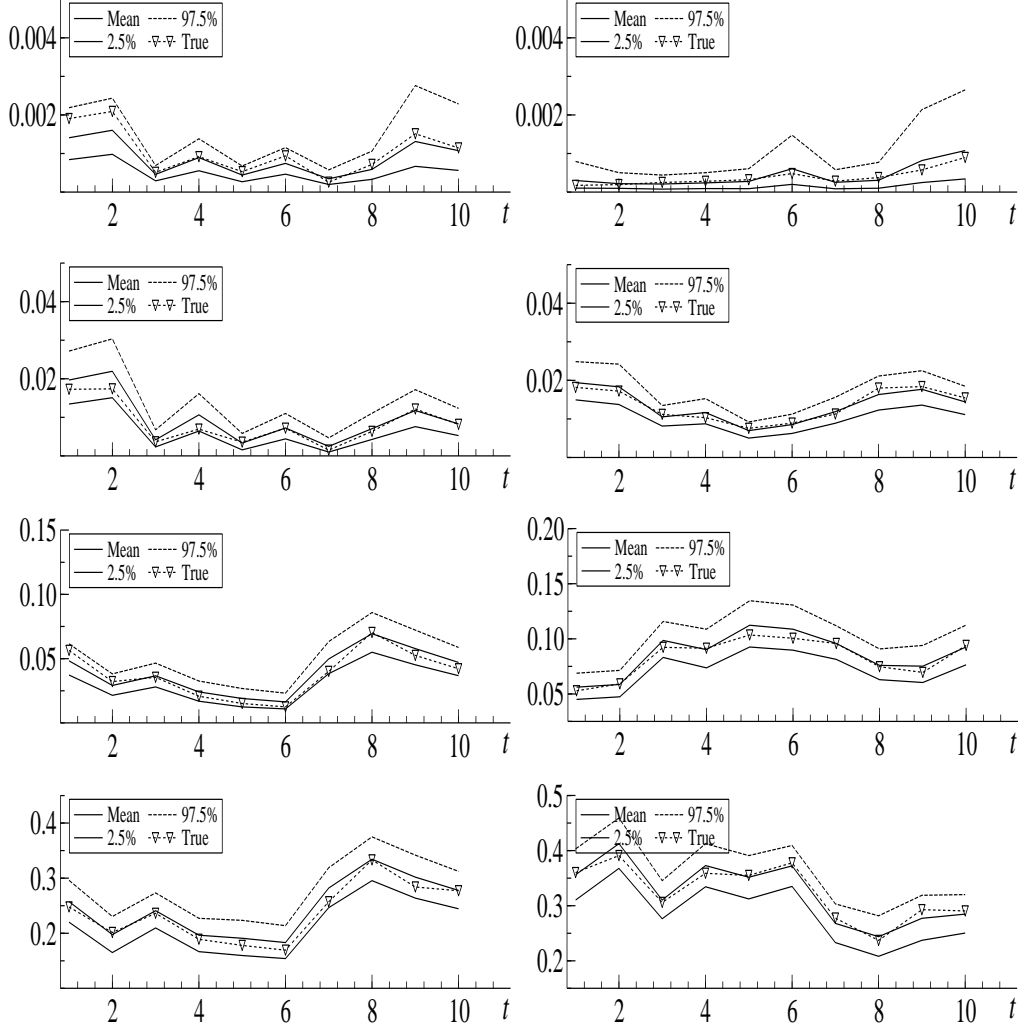
For the second component in (11),  $\sum_{h>k_t^+} \nu_{th} < k_t^+ \tilde{\epsilon}$  because  $\nu_{th} > \tilde{\nu}_{th} - \tilde{\epsilon}$  for  $h = 1, \dots, k_t^+$  and  $\sum_{h=1}^{k_t^+} \nu_{th} > 1 - k_t^+ \tilde{\epsilon}$ . In addition, it is straightforward to show that the proposed prior assigns positive probability to  $\mathcal{N}_{\epsilon''}(\boldsymbol{\psi}_{th}^{(j)})$ . Therefore, since  $D(\boldsymbol{\pi}^0, \epsilon)$  contains such case,  $\mathcal{Q}\{D(\boldsymbol{\pi}^0, \epsilon)\} > 0$ .

For any  $\boldsymbol{\pi} \in D(\boldsymbol{\pi}^0, \epsilon)$ ,

$$\begin{aligned}
\|\boldsymbol{\pi} - \boldsymbol{\pi}^0\| &= \sum_{t=1}^T p_t \sum_{c_1=1}^{d_1} \cdots \sum_{c_p=1}^{d_p} |\pi_{tc_1 \dots c_p} - \pi_{tc_1 \dots c_p}^0|, \\
&= \sum_{t=1}^T p_t \sum_{c_1=1}^{d_1} \cdots \sum_{c_p=1}^{d_p} \left| \sum_{h=1}^{\infty} \nu_{th} \prod_{j=1}^p \psi_{hc_j}^{(j)} - \sum_{l=1}^{k_t} \nu_{tl}^0 \prod_{j=1}^p \psi_{tlc_j}^{(j)} \right|, \\
&= \sum_{t=1}^T p_t \sum_{c_1=1}^{d_1} \cdots \sum_{c_p=1}^{d_p} \left| \sum_{h=1}^{k_t} \left( \nu_{tk_{t-1}^+ + h} \prod_{j=1}^p \psi_{k_{t-1}^+ + hc_j}^{(j)} - \nu_{th}^0 \prod_{j=1}^p \psi_{thc_j}^{(j)} \right) + \sum_{l \leq k_{t-1}^+, k_t^+ < l} \nu_{tl} \prod_{j=1}^p \psi_{l c_j}^{(j)} \right|, \\
&\leq \sum_{t=1}^T p_t \sum_{c_1=1}^{d_1} \cdots \sum_{c_p=1}^{d_p} \left( \sum_{h=1}^{k_t} \left| \nu_{tk_{t-1}^+ + h} \prod_{j=1}^p \psi_{k_{t-1}^+ + hc_j}^{(j)} - \nu_{th}^0 \prod_{j=1}^p \psi_{thc_j}^{(j)} \right| + \sum_{l \leq k_{t-1}^+, k_t^+ < l} \nu_{tl} \right), \\
&\leq \sum_{t=1}^T p_t \sum_{c_1=1}^{d_1} \cdots \sum_{c_p=1}^{d_p} \left( \sum_{h=1}^{k_t} \left| \nu_{tk_{t-1}^+ + h} - \nu_{th}^0 \right| + \sum_{l=1}^{k_t} \sum_{j=1}^p \left| \psi_{k_{t-1}^+ + lc_j}^{(j)} - \psi_{tlc_j}^{(j)} \right| + \sum_{l \leq k_{t-1}^+, k_t^+ < l} \nu_{tl} \right), \\
&= \sum_{c_1=1}^{d_1} \cdots \sum_{c_p=1}^{d_p} \sum_{t=1}^T p_t \sum_{h=1}^{\infty} |\nu_{th} - \tilde{\nu}_{th}| + \sum_{t=1}^T p_t \sum_{l=1}^{k_t} \sum_{j=1}^p \sum_{c_1=1}^{d_1} \cdots \sum_{c_p=1}^{d_p} \left| \psi_{k_{t-1}^+ + lc_j}^{(j)} - \psi_{tlc_j}^{(j)} \right|, \\
&< \prod_{j=1}^p d_j \epsilon' + \sum_{t=1}^T p_t k_t p \prod_{j=1}^p d_j \epsilon'', \\
&= \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{aligned}$$

Therefore  $\boldsymbol{\pi} \in \mathcal{N}_{\epsilon}(\boldsymbol{\pi}^0)$  and  $D(\boldsymbol{\pi}^0, \epsilon) \subset \mathcal{N}_{\epsilon}(\boldsymbol{\pi}^0)$ . Hence,  $\mathcal{Q}\{\mathcal{N}_{\epsilon}(\boldsymbol{\pi}^0)\} > 0$ .

## D Figures



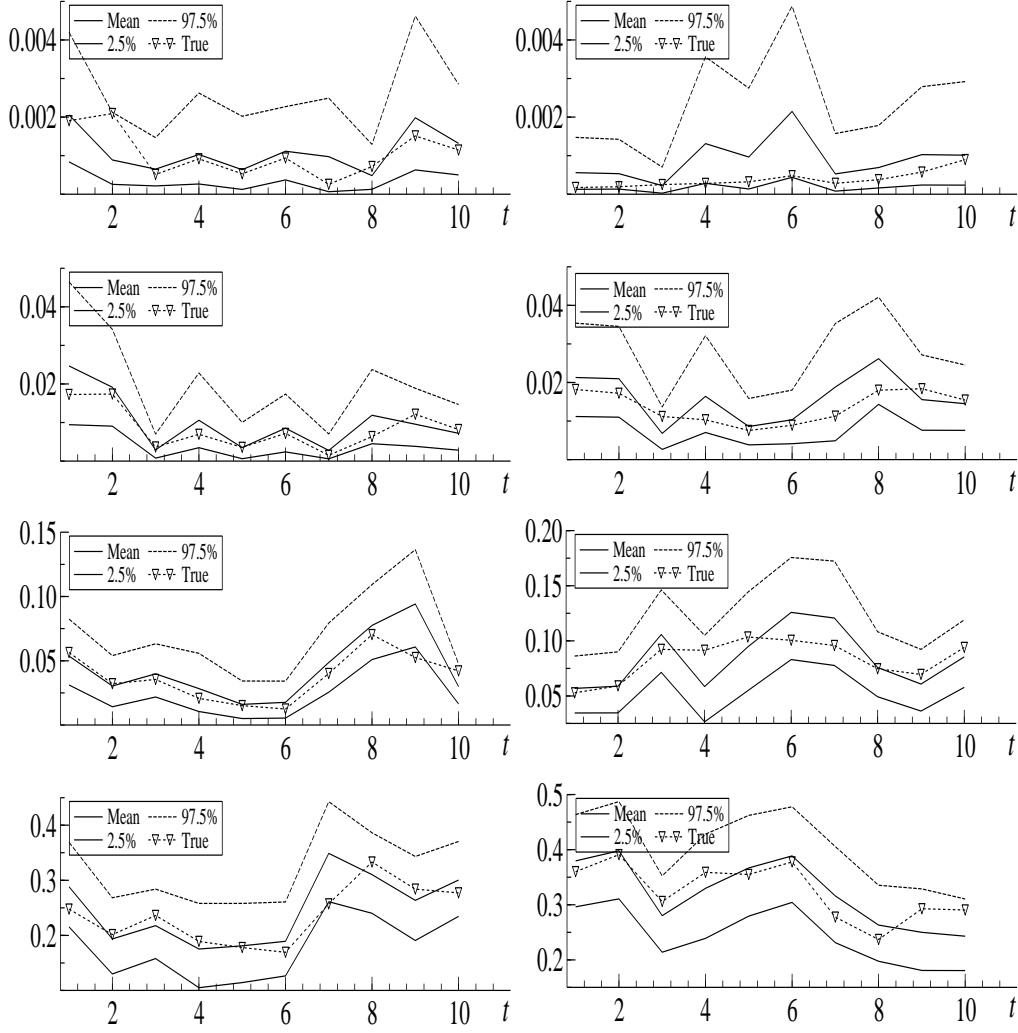
The first row:  $P(x_{ti4} = 0, x_{ti6} = 1, x_{ti10} = 2, x_{ti15} = 3)$  and  $P(x_{ti7} = 2, x_{ti9} = 0, x_{ti13} = 3, x_{ti19} = 1)$ .

The second row:  $P(x_{ti1} = 2, x_{ti7} = 1, x_{ti20} = 3)$  and  $P(x_{ti3} = 3, x_{ti12} = 1, x_{ti18} = 0)$ .

The third row:  $P(x_{ti11} = 1, x_{ti17} = 1)$  and  $P(x_{ti5} = 2, x_{ti19} = 1)$ .

The forth row:  $P(x_{ti8} = 0)$  and  $P(x_{ti20} = 3)$ .

Figure 1: Estimation results of cell probabilities by the proposed method.



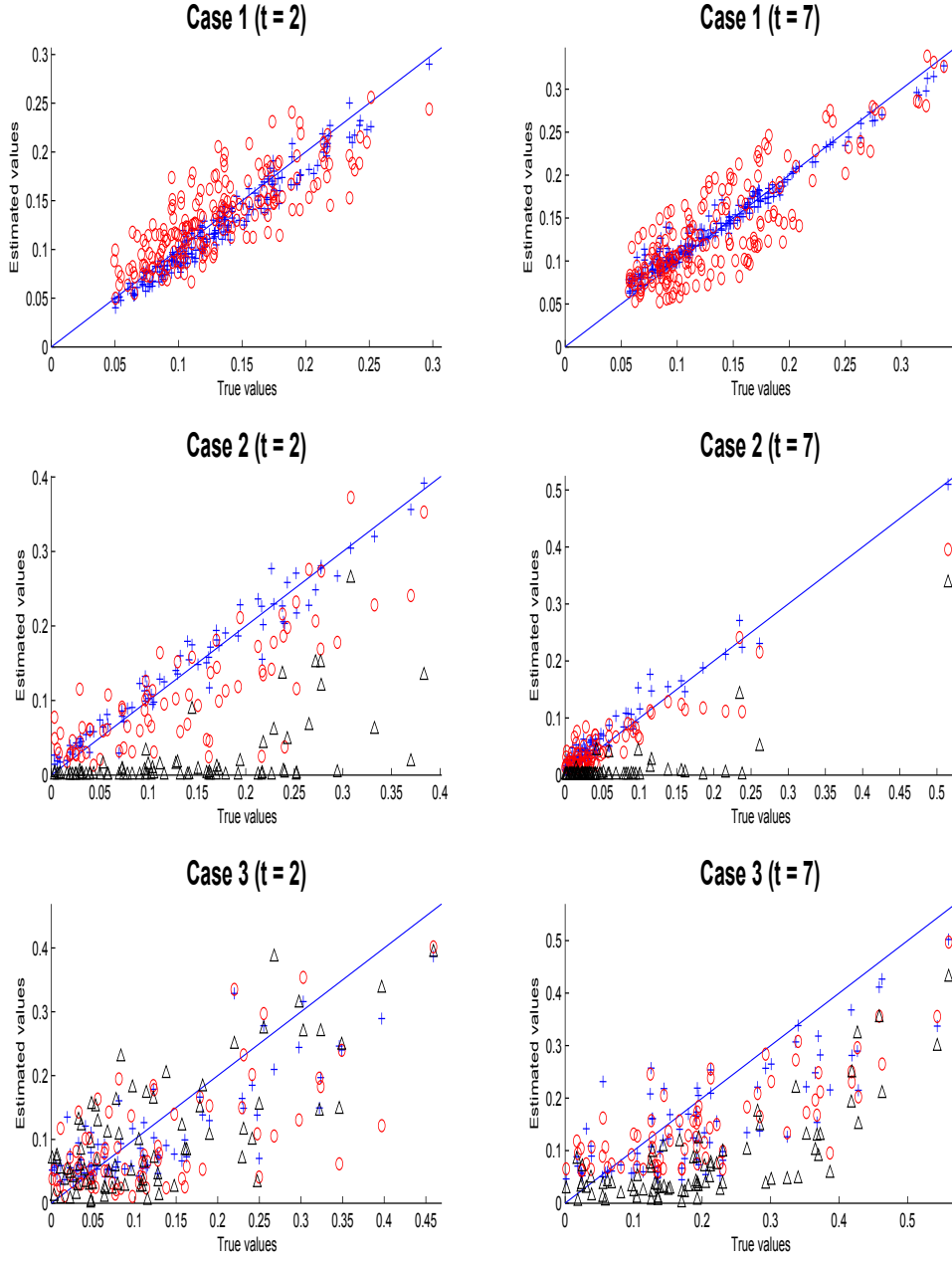
The first row:  $P(x_{ti4} = 0, x_{ti6} = 1, x_{ti10} = 2, x_{ti15} = 3)$  and  $P(x_{ti7} = 2, x_{ti9} = 0, x_{ti13} = 3, x_{ti19} = 1)$ .

The second row:  $P(x_{ti1} = 2, x_{ti7} = 1, x_{ti20} = 3)$  and  $P(x_{ti3} = 3, x_{ti12} = 1, x_{ti18} = 0)$ .

The third row:  $P(x_{ti11} = 1, x_{ti17} = 1)$  and  $P(x_{ti5} = 2, x_{ti19} = 1)$ .

The forth row:  $P(x_{ti8} = 0)$  and  $P(x_{ti20} = 3)$ .

Figure 2: Estimation results of cell probabilities by DX method.



$y$  axis represents estimated values and  $x$  axis true values. Cross-shaped dots represent the proposed method, circles DX method and triangles DH method. The first, second and third rows show the results at time  $t = 2$  and  $7$  using the first (case 1), second (case 2), third (case 3) simulation data sets.

Figure 3: Plots of true and estimated values of  $\rho_{tjj'}$  using the simulation data.



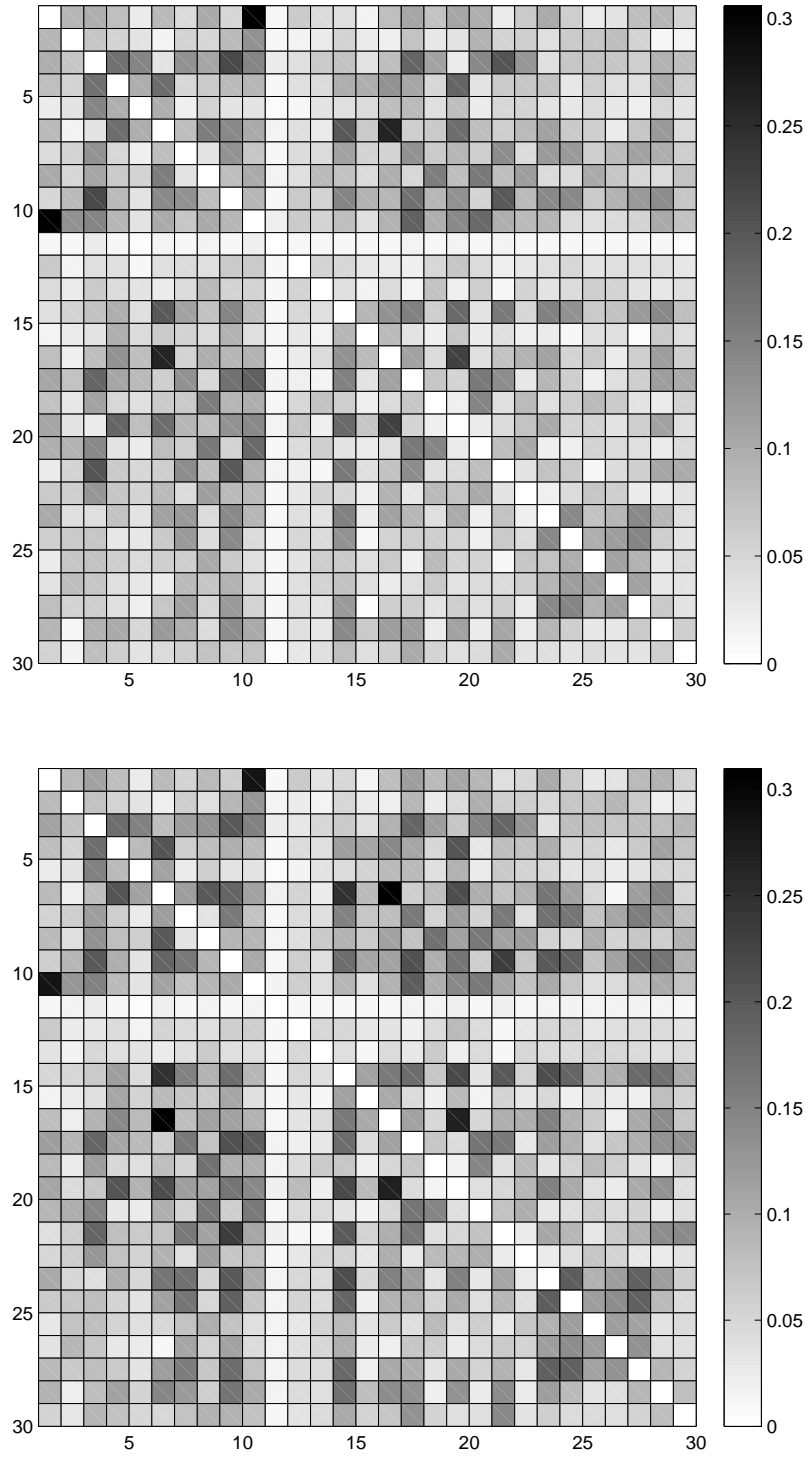
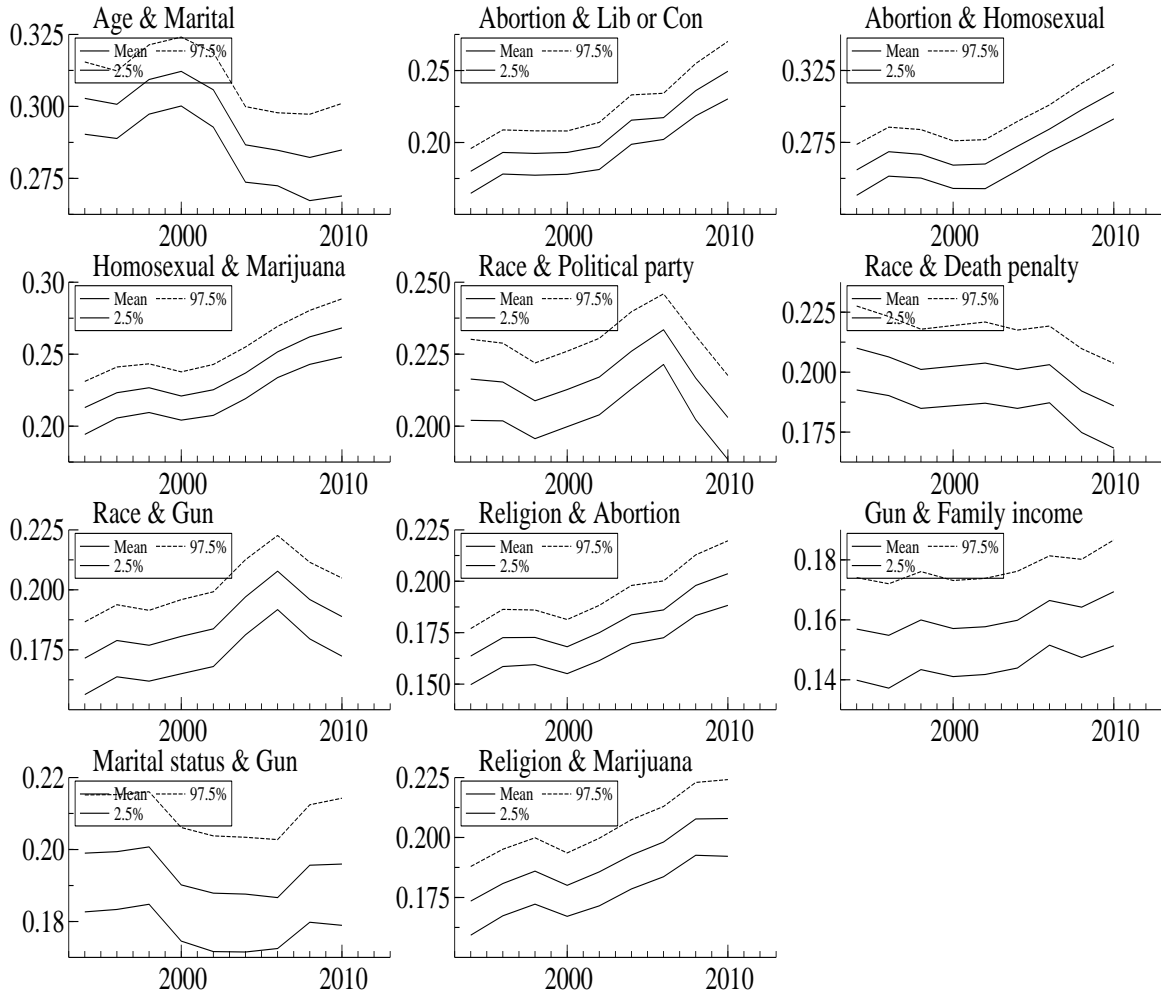


Figure 4: Posterior means of  $\rho_{tjj'}$  in 2002 (above) and 2010 (below).



The first row: (Age group, Current marital status), (Attitude toward abortion, Think of self as liberal or conservative) and (Attitude toward abortion, Attitude toward homosexual sex relations).

The second row: (Attitude toward homosexual sex relations, Should Marijuana be made legal), (Race, Political party affiliation) and (Race, Favor or oppose death penalty for murder).

The third row: (Race, Have gun in home), (Religious preference, Attitude toward abortion) and (Have gun in home, Total family income).

The fourth row: (Current marital status, Have gun in home) and (Religious preference, Should Marijuana be made legal).

Figure 5: Estimation results of  $\rho_{tjj'}$  for several pairs.

## References

Agresti, A. (2002). *Categorical Data Analysis*. (Second ed.). New York: Wiley.

Albert, J. H. and S. Chib (2001). Sequential ordinal modeling with applications to survival data. *Biometrics* 57, 829–836.

- Bhattacharya, A. and D. B. Dunson (2011). Simplex factor models for multivariate unordered categorical data. *Journal of the American Statistical Association*. to appear.
- Bush, C. and S. MacEachern (1996). A semiparametric bayesian model for randomised block designs. *Biometrika* 83, 275–285.
- Carter, C. K. and R. Kohn (1994). On gibbs sampling for state space models. *Biometrika* 81, 541–553.
- Chung, Y. and D. B. Dunson (2009). Nonparametric bayes conditional distribution modeling with variable selection. *Journal of the American Statistical Association* 104, 1646–1660.
- Chung, Y. and D. B. Dunson (2011). The local dirichlet process. *Annals of the Institute for Statistical Mathematics* 63, 59–80.
- Dahinden, C., M. Kalisch, and P. Buehlmann (2010). Decomposition and model selection for large contingency tables. *Biometrical Journal* 52, 233–252.
- Dahinden, C., G. Parmigiani, M. Emerick, and P. Buehlmann (2007). Penalized likelihood for sparse contingency tables with an application to full-length cDNA libraries. *BMC Bioinformatics* 8.
- de Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika* 82, 339–350.
- Dobra, A. and A. Lenkoski (2011). Copula gaussian graphical models and their application to modeling functional disability data. *Annals of Applied Statistics* 5, 969–993.
- Doornik, J. (2006). *Ox: Object Oriented Matrix Programming*. London: Timberlake Consultants Press.
- Dunson, D. B. (2006). Bayesian dynamic modeling of latent trait distributions. *Biostatistics* 7, 551–568.
- Dunson, D. B. and C. Xing (2009). Nonparametric bayes modeling of multivariate categorical data. *Journal of the American Statistical Association* 104, 1042–1051.
- Durbin, J. and S. J. Koopman (2002). Simple and efficient simulation smoother for state space time series analysis. *Biometrika* 89, 603–616.
- Fienberg, S. and A. Rinaldo (2007). Three centuries of categorical data analysis: Log-linear models and maximum likelihood estimation. *Journal of Statistical Planning and Inference* 137, 3430–3445.

- Früwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis* 15, 183–202.
- Ghosh, J. and R. Ramamoorthi (2003). *Bayesian Nonparametrics*. Springer Verlag.
- Griffin, J. E. and M. F. J. Steel (2006). Order-based dependent dirichlet processes. *Journal of the American Statistical Association* 101, 179194.
- Harshman, R. A. (1970). Foundations of the parafac procedure: Models and conditions for an "explanatory" multimodal factor analysis. *UCLA Working Papers in Phonetics* 16, 1–84.
- Ishwaran, H. and L. F. James (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* 96, 161–173.
- Kalli, M., J. E. Griffin, and S. G. Walker (2011). Slice sampling mixture models. *Statistics and Computing* 65, 93105.
- Kolda, T. G. (2001). Orthogonal tensor decompositions. *SIAM Journal on Matrix Analysis and Applications* 23. (in press).
- MacEachern, S. N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science, Alexandria, VA: American Statistical Association*, 50–55.
- MacEachern, S. N. (2000). Dependent dirichlet processes. Technical report, Ohio State University, Department of Statistics.
- Papaspiliopoulos, O. and G. O. Roberts (2008). Retrospective markov chain monte carlo methods for dirichlet process hierarchical models. *Biometrika* 95, 169186.
- Ren, L., D. Dunson, S. Lindroth, and L. Carin (2010). Dynamic nonparametric bayesian models for analysis of music. *Journal of the American Statistical Association* 105, 458–472.
- Rodriguez, A. and D. B. Dunson (2011). Nonparametric bayesian models through probit stick-breaking processes. *Bayesian Analysis* 6, 145–177.
- Rodriguez, A. and E. T. Horst (2008). Bayesian dynamic density estimation. *Bayesian Analysis* 3, 339366.
- Sethuraman, J. (1994). A constructive definition of dirichlet priors. *Statistica Sinica* 4, 639–650.

Walker, S. G. (2007). Sampling the dirichlet mixture model with slices. *Communications in Statistics-Simulation and Computation* 36, 45–54.

## E Supplemental materials

No.	Categorical variable (Name in GSS)	# of categories
1	Age group* (AGE)	8
2	Sex (SEX)	2
3	Race (RACE)	3
4	Religious preference** (RELIG)	5
5	Region (REGION)	9
6	Attitude toward abortion (ABANY)	2
7	Should Govetnment help pay for medical care? (HELPSICK)	5
8	Highest degree (DEGREE)	5
9	Political party affiliation (PARTYID)	8
10	Current marital status (MARITAL)	5
11	Astrological sign (ZODIAC)	12
12	Confidence in banks and financial institutions (CONFINAN)	3
13	Confidence in U.S. Supreme Court (CONJUDGE)	3
14	Think of self as liberal or conservative (POLVIEWS)	7
15	Belief in life after death (POSTLIFE)	2
16	Attitude toward homosexual sex relations (HOMOSEX)	5
17	Have gun in home (OWNGUN)	2
18	Subjective class identification (CLASS)	4
19	Should Marijuana be made legal (GRASS)	2
20	Total family income (INCOME)	12
21	Favor or oppose death penalty for murder (CAPPUN)	2
22	Attitude toward spending money on space exploration program (NATSPAC)	3
23	Attitude toward spending money on improving and protecting environment (NATENVIR)	3
24	Attitude toward spending money on improving and protecting the nations's health (NATHEAL)	3
25	Attitude toward spending money on halting the rising crime rate (NATCRIME)	3
26	Attitude toward spending money on dealing with drug addiction (NATDRUG)	3
27	Attitude toward spending money on improving the nation's education system (NATEDUC)	3
28	Attitude toward spending money on the military, armaments and defense (NATARMS)	3
29	Attitude toward spending money on foreigh aid (NATAID)	3

\*The category of Age group is different from the original one: 1. 18 or 19 years old, 2. 20s, 3. 30s, 4. 40s, 5. 50s, 6. 60s, 7. 70s, 8. more than 80 years old.

\*\*The category of Religious preference is different from the original one: 1. Protestant, 2. Catholic, 3. Jewish, 4. None, 5. Others.

Table 7: List of categorical variables.

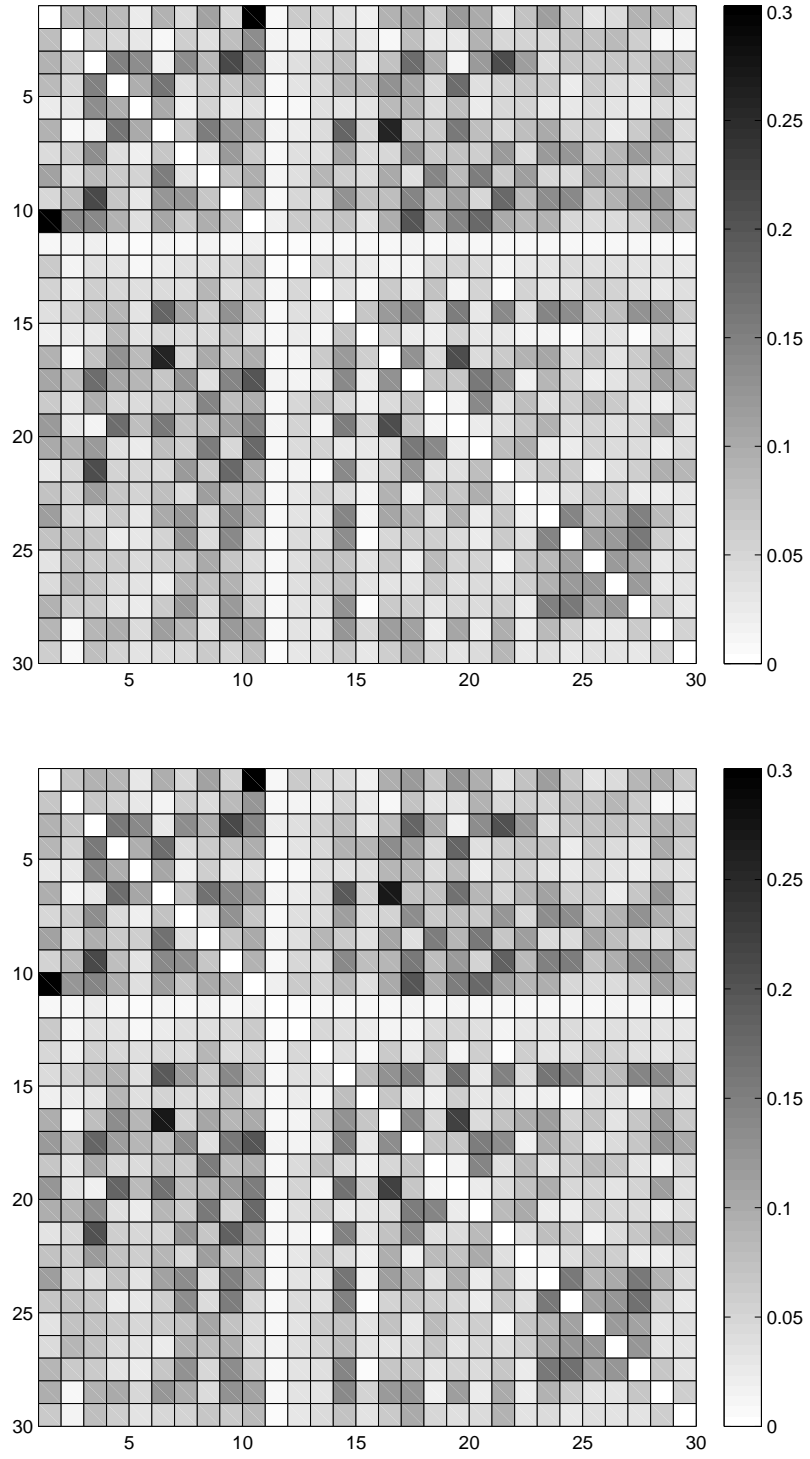


Figure 6: Posterior means of  $\rho_{tjj'}$  in 1994 (above) and 1996 (below).

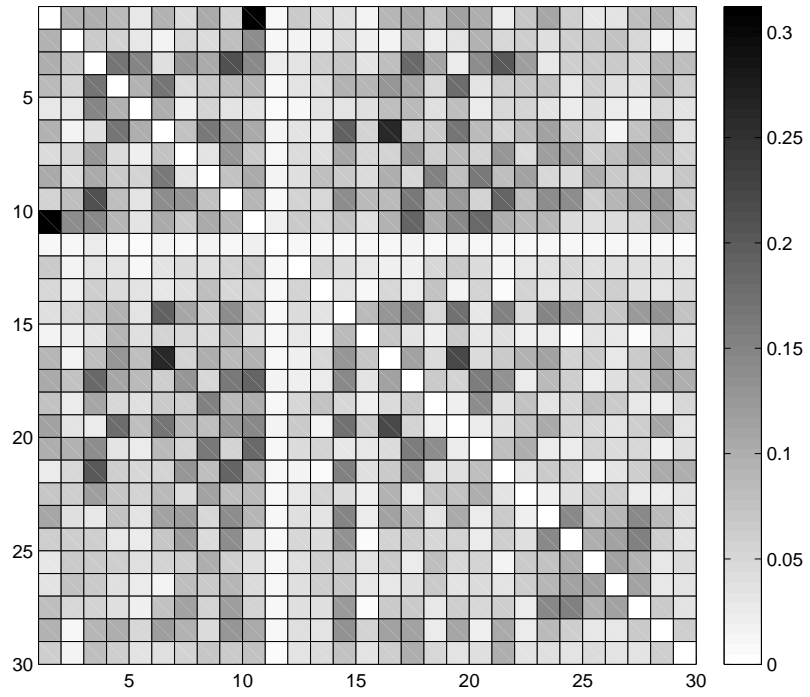
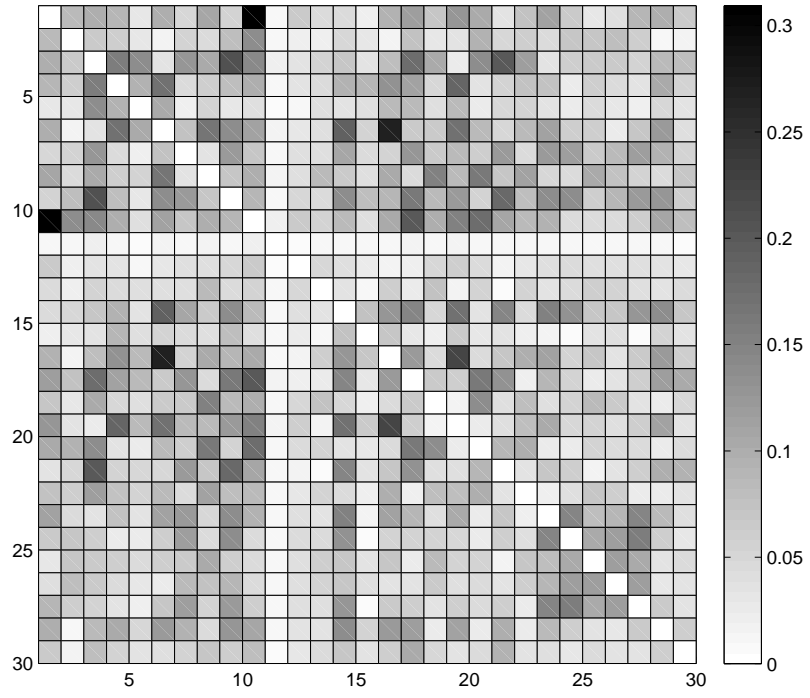


Figure 7: Posterior means of  $\rho_{tjj'}$  in 1998 (above) and 2000 (below).



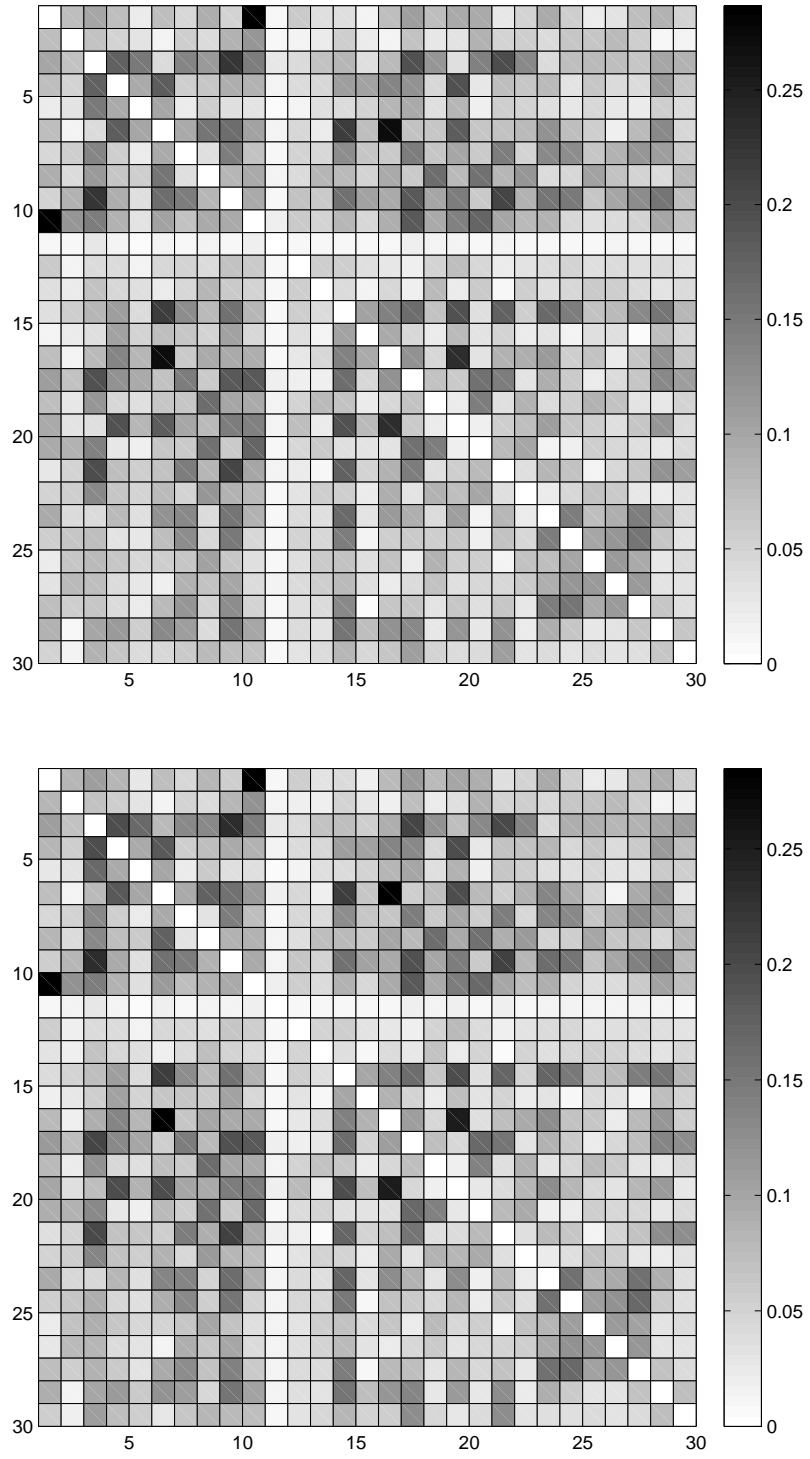


Figure 8: Posterior means of  $\rho_{tjj'}$  in 2004 (above) and 2006 (below).

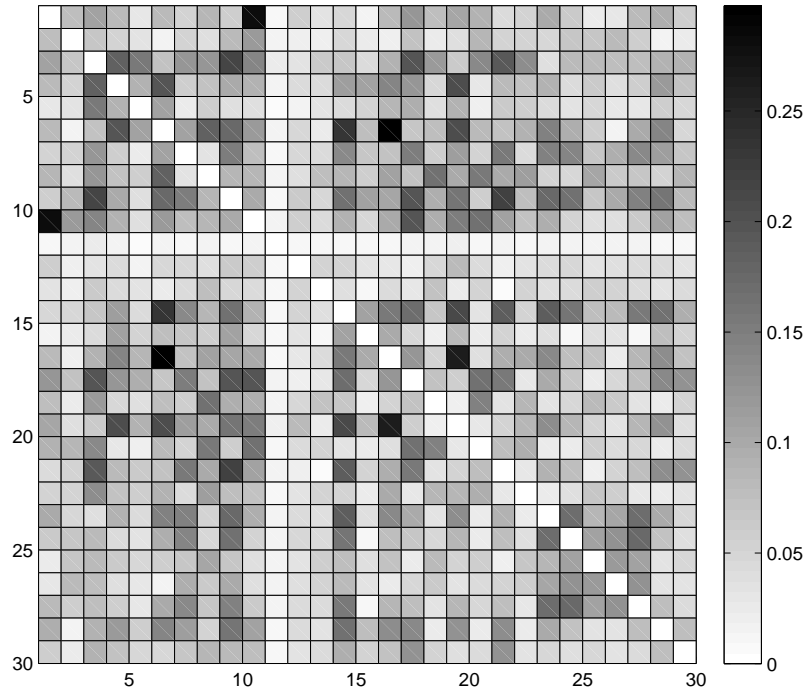


Figure 9: Posterior means of  $\rho_{tjj'}$  in 2008.